对『伪心理学』说不

第 8 版

【加】基思·斯坦诺维奇 著
窦东徽 刘肖岑 译

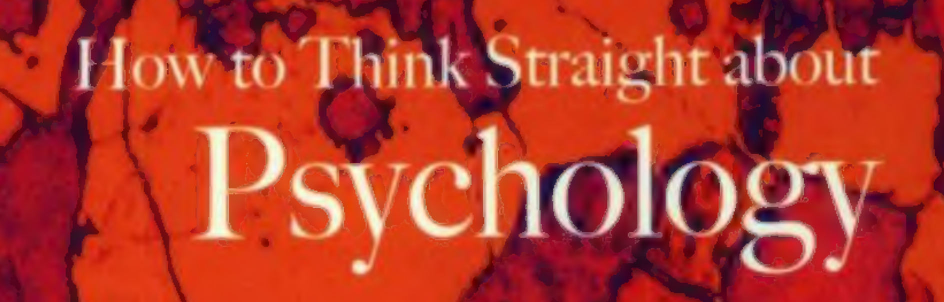
8th edition

How to Thin[k]

Psy[c]

KEITH E. STANOVICH

人民邮电出版社
POSTS & TELECOM PRESS



在浩如烟海、良莠不齐的心理学信息面前，如何拨除迷雾，去伪存真，成为一个明智的心理学信息的消费者？这本书将教给你科学实用的批判性思维技能，将真正的心理学研究从伪心理学中区分出来，告诉你什么才是真正的心理学。

这本“与众不同的心理学”教科书，总结了心理学家的职业特质，让每一位读者都有机会去理解我们心理学家是如何去思考、分析和解读人类的行为和心理的。每一章都将一个常识的、朴素的、直觉的有关人类心理的分析和思考与一个科学的、严谨的、心理学的分析和思考相对比，以帮助读者理解心理学家的分析逻辑和研究思路。

——彭凯平教授　清华大学心理学系系主任
美国加州大学伯克利分校心理学系终身教授

这本书在第一版时，我已经开始把它推荐给我的学生，那是十多年前的事了。现在它已经出到第八版了，可见它深受几代老师及学生的喜爱。阅读本书，可以让人们认识到及分辨出什么样的研究及数据才是真正科学的及可信的。为此，我不单只把它推荐给大专院校心理学系的学生作为研究方法课的教材，也强烈推荐给一般读者。

——杨中芳教授　中山大学心理学系前系主任

北京新曲线出版咨询有限公司
网址：www.ncc-pub.com



人民邮电出版社 POSTS & TELECOM PRESS

PEARSON
ALWAYS LEARNING

PEARSON

分类建议：心理学
人民邮电出版社网址：www.ptpress.com.cn
www.pearson.com

对“伪心理学”说不

第8版

[加]基思·斯坦诺维奇 著

窦东徽 刘肖岑 译

人民邮电出版社

北 京

图书在版编目（CIP）数据

对“伪心理学”说不：第8版 /（加）斯坦诺维奇（Stanovich，K.E.）著；窦东徽，刘肖岑译．-北京：人民邮电出版社，2012.1（2012.1 重印）

ISBN 978-7-115-26325-4

Ⅰ．①对…　Ⅱ．①斯…②窦…③刘…　Ⅲ．①心理学　Ⅳ．①B84

中国版本图书馆 CIP 数据核字（2011）第 189637 号

Keith E. Stanovich

How to Think Straight about Psychology, 8th Edition

ISBN 0-205-48513-8

Posts & Telecom Press is authorized by Pearson Education to publish and distribute exclusively this reprint edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong SAR, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Copyright © 2012 by Pearson Education, Inc. and Posts & Telecom Press.

本书中文简体字版由 Pearson Education 公司授权人民邮电出版社独家出版发行。此版本仅限在中华人民共和国境内（不包括中国香港、澳门特别行政区及中国台湾地区）销售。未经授权的本书出口将被视为违反版权法的行为。

未经出版者预先书面许可，不得以任何方式复制或发行本书的任何部分。

本书封底贴有人民邮电出版社和 Pearson Education 公司防伪标签，无标签者不得销售。

北京市版权局著作权合同登记号：01-2008-1156

版权所有，侵权必究。

对“伪心理学”说不（第8版）

◆ 著　[加] 基思・斯坦诺维奇
译　窦东徽　刘肖岑
策　划　刘　力　陆　瑜
责任编辑　刘冰云
装帧设计　陶建胜

◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街14号A座
邮编　100061　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
电话　（编辑部）010-84937150　（市场部）010-84937152
（教师服务中心）010-84931276
三河市李旗庄少明印装厂印刷
新华书店经销

◆ 开本：880×1230　1/32
印张：9
字数：250千字　2012年1月第1版　2012年1月第2次印刷
著作权合同登记号　图字：01-2008-1156

ISBN 978-7-115-26325-4/F

定价：28.00元

本书如有印装质量问题，请与本社联系　电话：(010) 84937153

内容提要

在今天的大众媒体和图书市场上，到处充斥着关于潜能提升、心理操控、色彩星座、催眠读心等伪装成心理学的主题，更有一些伪心理学家、所谓的心理治疗师打着心理学的旗号欺世盗名，从中渔利。在浩如烟海、良莠不齐的心理学信息面前，如何拨除迷雾，去伪存真，成为一个明智的心理学信息的消费者呢？这本书将教给你科学实用的批判性思维技能，将真正的心理学研究从伪心理学中区分出来，告诉你什么才是真正的心理学。

本书第1版出版于1983年，20多年来一直被奉为心理学入门经典，在全球顶尖大学中享有盛誉，现在呈现在读者面前的是第8版。这本书并不同于一般的心理学导论类教材，很多内容是心理学课堂上不曾讲授的，也是许多心理学教师在教学中感到只可意会而不可言传的。作者正是从此初衷出发，以幽默生动的语言，结合一些妙趣横生、贴近生活的实例，深入浅出地介绍了可证伪性、操作主义、实证主义、安慰剂效应、相关和因果、概率推理等心理学中的基本原则。与上一版相比，第8版更新了最新的研究资料和实例，扩展了信度、效度、元分析、随机取样与随机分配等问题的讨论。

本书不仅适合于心理学专业的学生，有助于建立心理学研究中必要的批判性思维技能与意识，而其通俗易读性也非常适合所有对心理学感兴趣的读者，它将帮助你纠正对心理学的种种误解，学会独立地评估心理学信息，用科学的精神和方法理解自己和他人的行为。此外，由于心理学与其他学科的共通性，本书也不失为一本精彩有趣的科学哲学类读物。

心理学有什么与众不同之处

心理学是一门很容易让人误解的学科。与其他学科不同，它研究的是与人民生活紧密相关的问题。有些还正好是大家都熟悉，而且经常关心的问题。比如：什么样的人容易讨人喜欢？什么样的事情我们记忆深刻？什么样的事情让人高兴（或痛苦）？为什么男人比女人更爱聊政治时事？人为什么要自杀？意念能不能被植入梦中？等等……正因为大家关心这些问题，人们就会有自己的分析，自己的证据，得出自己相信的结论。很多时候，这些自觉的结论与心理学家的研究结论并不完全一致，比如说，我们心理学家就发现，青梅竹马的婚姻很难成立，婴儿并不是有奶便认娘，性格并不决定一个人的命运，等等……每当矛盾、怀疑、迷惑甚至气愤产生的时候，我们到底是该相信自己的直觉、经验和常识呢，还是该相信心理学的研究、证据和知识呢？

不幸的是，大多数的心理学教科书只满足告诉大家心理学的研究、发现和知识，但从不说明为什么这些研究、发现和知识是值得我们关注和信任的。这些书的作者好像都在假设每一位读者思考问题起来都像心理学家一样，相信和理解心理学的研究、发现和知识。幸运的是，Keith Stanovich 教授写了这本 How to Think Straight about Psychology，一本“与众不同的心理学”教科书。他总结了心理学家的职业特质，让每一位读者都有机会去理解我们心理学家是如何去思考、分析和解读人类的行为和心理的。他的每一章都将一个常识的、朴素的、直觉的有关

人类心理的分析和思考与一个科学的、严谨的、心理学的分析和思考相对比，以帮助读者理解心理学家的分析逻辑和研究思路。

我个人觉得，在 Stanovich 阐述的心理学特质中，有两点应该是区分我们心理学家和其他人（包括其他领域的科学家）与众不同的地方。

其一是我们心理学的批判性思维习惯，也就是说，心理学不相信个人的智慧，更相信科学的方法，而科学方法的本质是证伪，即对我们的经验、常识和直觉，产生怀疑、挑战和批评。从原则上而言，心理学家不怕犯错误，但害怕以假乱真。心理学家也不相信那些能回答所有问题的绝对真理，但相信对所有问题应该有一个相对正确的答案。心理学家从不相信个案和例子，因为其随机性和主观性太过明显，但我们愿意相信大样本基础上的科学研究发现。我们希望听到动听的心理故事，但更愿意看到众多心理学观察的数据和总结。

其二是我们心理学的概率性思维习惯。我们和很多自然科学家思考方式不同之处就在于我们更容易相信，任何人类的行为都是概率性的表现，也就是说，它有一定的不确定性，会受到其他随机事件的影响。其实人类很多学科都是建立在概率基础之上的，比如说核物理、天体物理、生物进化、病理学、所有的社会科学（经济学、社会学、政治学等）等等，它们都不可能准确预测每一个研究对象的具体活动，但都对整体的事物活动规律有很好的描述和预测。只不过我们心理学家更愿意承认而且强调我们学科的不确定性而已。

总之，我很高兴地看到这本书在中国再版。杨中芳老师的早期译本是我很喜欢向心理学爱好者推荐的一本心理学教科书。而新版的译本，尤其令我兴奋，因为它是由我欣赏的两位年轻同事，东黴和肖岑，在清华大学的青年教师公寓里，下工夫，花时间，认真翻译完成的。两位既是同行，也是夫妻，应该是我们心理学界不鲜见的学术伴侣之一。这可能也是我们心理学界与众不同的地方吧。

彭凯平

清华大学心理学系教授、系主任

译者序

每个学心理学专业的人都会被人问到三个问题，第一个通常是："你知道我现在心里在想什么吗？"名列榜眼的问题可能是："那你一定会催眠和解梦喽？"排名第三的问题则一般会是："什么，你们心理学还做实验？"这类对于问话人来说再正常不过的问题，却总能让这些学心理学的人哭笑不得。

但是，一个真正称职的心理学工作者是不应回避或敷衍这类问题的，正确的做法是直面它们并认真分析其中的意义，给予人们正确的解答。从最常被问到的这三个问题中，我们便能以管窥豹地看出公众对于心理学的一些常见的误解和迷思。

第一个问题所揭示的事实是，心理学在大众心目中被万能化了。所谓万能化，是指心理学研究对象和功能的边界被无限地夸大了。首先，许多人认为心理学无所不包，星座、属相、血型、玄学、人生哲理及各种心灵鸡汤都被认为是心理学的范畴——实际上，心理学研究虽然涉及广泛的人的客观行为和心理现象，但作为心理学的研究对象必须满足"可检验"这一标准，超验的、超感官的问题不在科学心理学的研究之列；另外，并不是说心理学不能研究血型、星座和超感知，实际情况是，这类现象和理论在历史上都曾经是心理学的研究对象，但已被各种科学的方法所证伪并从心理学的研究对象中剔除了出去。还有一种观点认为，心理学是无所不能的，学了心理学就能读心、能算命，甚至能千里之外控制他人的大脑并操纵其行为，还兼具其他种种神奇的功效。我们说心理学很神奇，是因为它能在一定程度上解释心理现象和预测行为，并能

提供一些行之有效的干预方法，但是，科学心理学所有的预测和解释都基于客观数据，所有的结论都具有概率的性质，有一定的适用范围和条件，干预方案也必须符合客观的心理和行为规律。因此没有能够完美解释一切现象的心理学理论，也没有超出人类经验以外的干预和应用。

第二个关于催眠的问题则揭示了另外一种对于心理学常见的误解：心理学就是弗洛伊德的精神分析，就是心理咨询。弗洛伊德这位伟大的心理学家确实让后世的心理学研究者“既爱又恨”。弗洛伊德的功绩在于，他以其非凡的工作投入和严谨的思辨构筑了一套宏大、晦涩而让人着迷的理论体系，开启了一片探索人类心灵世界的全新领域，对社会文化、艺术创作产生了不可估量的影响，并真正让世人开始了解和重视心理学；迄今为止他的理论仍然广泛应用于心理治疗和干预的诸多领域。今天人们都能熟练用“焦虑”、“压抑”、“潜意识”这些词来描述或解释分析自己的心理状态，也是拜弗洛伊德和他的精神分析学派所赐。说到心理学家，或许很多人不知道冯特和斯金纳，但没有人会不知道弗洛伊德。然而，弗洛伊德的盛名也带来了一系列问题，最主要的是两点：一是他的光芒掩盖了其他领域心理学研究者的工作和成就，二是弗洛伊德所构建的这些精巧的理论很难用实证的方法加以验证，这也为后来众多伪心理学和虚假治疗手段的滋生和发展留下了一道后门。

最后一个问题也是最关键的——公众不了解主流的科学心理学已经在实证主义的道路上行进了很远。心理学是一门科学，必然要遵循科学的标准，即研究的必须是实证可解的问题；方法上要遵循系统的实证主义；研究结论要能够被重复验证，并能经过同行评议获得认可。这三条标准也是区分科学心理学和其他伪心理学的分水岭。心理学的研究因此也与操作定义、实验组和控制组、变量控制、数据统计等词汇联系起来，而不是个人化的体验及感悟、头脑风暴或纯粹的逻辑推导。

那么，是什么阻碍了公众了解和认知真正的科学心理学？原因应该是多方面的。这其中既有大众的原因，也有学院派研究者自身的问题，当然也包括学科自身特点的限制。

首先，对心理学的渴求使得大众对心理学产生了特殊的期待。心理现象和心理问题与每个人休戚相关，人们试图通过心理学解释这些问题和现象，并从中获得行之有效的帮助和建议。在这方面，针对个体的精神分析和治疗技术有一定的优势，但其他大多数心理学研究则针对的是群体的普遍行为规律，偏重于解释和预测，其研究结论都具有概率性和领域特异性，因此无法有针对性地、面面俱到地解决个体所有的心理问题。这有时难免会让一部分抱着“求医问药”的心态来了解心理学的人感到失望。

第二，术语体系产生的阻隔。心理学作为科学，为了让研究者之间形成共识并利于重复验证，产生了一套严谨的术语体系，这一点上和其他科学学科并无二致。但不同的是，没有人会因为不懂量子力学的各种晦涩的术语而鄙视物理学，但他们却会因不能忍受心理学术语所造成的隔膜而对科学心理学产生拒斥，继而转向了那些通俗易懂但科学性无法得到保证的理论或方法。究其原因，还是因为心理问题和现象与每个人的关系太为紧密了，最近的需要总是要以最快捷的方式介入和解决，这就构成了快速实用和科学研究之间的一对矛盾。

第三，心理学许多基础研究确实和大众生活有一定的距离。首先是因为这些研究往往是控制十分严密的实验室实验，探讨的是最基本的生理、行为和机制，有些唤起特定反应的刺激可能在日常生活中几乎不可能出现（如人造的假字等等）；其次，并不是说基础研究的成果没有任何实用价值，而是很多成果转化为实际的应用需要一段时间，有的研究发现甚至要几十年之后才能体现出其应用价值。这种滞后性有时会让人产生一种科学心理学研究毫无意义的错觉。

第四，许多学院派心理学工作者的做法也难辞其咎。首先，虽然说许多基础研究和大众生活有一定的距离，但心理学的研究的最终目的是要服务于人类。基础研究之外还有许多与现实生活紧密相关的理论及应用研究。这方面好的研究应当是“现象为师”的，即从社会生活现象出发去发现问题、做出假设、收集观测数据、进行验证，最终得出的结论

也应该能够在一定程度上解释或解决现实问题。但许多研究者往往是从概念出发，空对空地进行一些干瘪乏味的研究，结论的实用性和推广性十分有限；另外，学院派的心理学工作者普遍的一个问题是，醉心于研究而不注重或不善于将成果进行普及和推广，这块阵地最终被伪心理学家和各种江湖术士所攫取。

最后一个原因，大众媒体在科学心理学的传播方面做得并不够好。随着心理学的热度持续升温，许多电视、电台、网站和出版商和纸媒也不断推出与心理学相关的节目、书籍或专栏，但由于种种原因（如专业限制、商业考虑等等），最终呈现在大众面前并得到广泛传播的往往并不是科学的心理学，而是包装精美的伪心理学；有一些所谓的“心理学家”甚至在媒体平台上用错误的理论在误导大众。与之相对应的是，真正的心理学家和专业书籍无人问津，而星座、血型、养生以及各种未经实证检验的古怪疗法却打着心理学的旗号招摇过市，让心理学蒙受了许多质疑和指责。

基于以上种种，如何让公众了解真正科学的心理学变得十分必要。市场上有关心理学的书籍很多，但对象分化的情况也很突出：针对心理学专业学生和心理学工作者的专业书籍能够提供很多有用的信息和知识，但有专业门槛的隐形限制；而针对普通读者的非专业书籍相对通俗但在信息质量方面良莠不齐。Keith E. Stanovich 教授所写的这本《对“伪心理学”说不》的定位则兼顾心理学的初学者和对心理学感兴趣的更广大的读者群体，从质量来说应为此类心理学入门读物中之翘楚。在这本 20 万字左右的小册子里，作者以生动而严谨的笔触告诉了读者什么才是真正的心理学。书中重申了科学心理学诸多重要标准和核心理念，澄清了有关心理学的种种误解和迷思，有破有立，言之凿凿；理论讲述与精彩的实验案例交相呼应，集科学性和趣味性于一体，十分耐读；有些犀利的论点足以对读者原有的知识信念构成挑战，但这种不安很快就会被知识重构的提升感和思辨的乐趣所取代。

这本书先前的版本曾经由杨中芳老师翻译，已使很多读者从中受益。

第 8 版则由新曲线出版咨询公司委托我与刘肖岑老师共同翻译，接到任务深感责任重大。翻译的过程对我们来说也是一个学习和审视自己的过程。书已近付梓，但译者水平所限，难免有不周或纰缪之处，还请广大读者给予指正。

最后我们要感谢北京师范大学心理学院邹泓教授、华东师范大学心理系桑标教授以及美国加州大学伯克利分校教授、清华大学心理学系教授彭凯平（Kaiping Peng）在本书翻译过程中给予的指点和帮助，以及一直以来在学术和思想方面的引导和教诲。同时这本书的顺利出版，与新曲线出版公司刘力先生的信任，以及刘丽丽和刘冰云两位编辑认真负责的工作是分不开的，在此一并致谢。

美国心理学会前任主席、积极心理学创始人马丁·塞利格曼教授曾用一个、两个和三个词形容心理学现在的状态，分别是："good"（好）、"not good"（不好）和"not good enough"（还不够好）。确实，心理学是一门很好的科学，它有着辉煌的过去、蓬勃的现在和光明的未来；虽然也有着不尽如人意的地方，但我们坚信，现在所做的一切工作，都是在让心理学变得更好。

窦东徽　刘肖岑

2011 年盛夏于清华大学青年公寓

To Paula, who taught me how to think straight about life

序 言

有这样一门尚不为大多数人所知晓的知识，它涉及人类行为和意识的不同形式，可以用来解释、预测和控制人类的行为。与没有掌握这门知识的人相比，学习过这门知识的人能够更好地理解他人，更加全面和精确地认识到哪些原因决定他人的想法和行为。

你可能想不到，这门不被知晓的知识就是心理学。

当我说心理学仍不为人所知时意味着什么？你一定认为此话不能当真。如今，书店里充斥着大量所谓心理学的书籍，电视和广播脱口秀定时播放关于心理学的主题，报纸和杂志上也常有所谓的心理学家对各种问题发表意见，怎么能说心理学无人知晓呢？但从某种关键的意义上来讲，心理学确实仍是一个不为人知的知识领域。

尽管心理学似乎得到了众多媒体的关注，但是心理学这个知识体系的绝大部分内容仍不为公众所知。经由大众媒体传播的“心理学”知识在很大程度上只是一种幻象。很多人不知道他们在许多书店里看到的大部分所谓的心理学书籍，都是由一些在心理学界根本站不住脚的人写的；很多人也不知道，多数在电视上号称心理学家的人，根本得不到美国心理学会（APA）和美国心理协会（APS）的承认；他们更不知道，大多数频频亮相的心理学“专家”，其实对心理学领域的知识积累没有做出丝毫的贡献。

媒体这种对于“心理学”话题的浅薄关注，不仅向公众传递了许多错误信息，还遮蔽了心理学领域中真正的、不断发展的知识。公众不能肯定到底哪些是心理学，哪些又不是，也不知道该怎样独立地对有关人类行为的主张做出判断。更大的问题在于，很多人始终觊觎着那些要么

缺乏判断能力、要么认为无法对心理学言论做出评估的公众。后一种观点有时被称为“怎么着都行”的态度，是本书要讨论的谬误之一，这种心态对社会危害极大。公众不知道，关于行为的言论是可以验证的，很多伪心理学正是利用公众的这种无知建立起百万美元的产业。人们并不知道许多伪心理学（例如占星术、通灵外科手术、超速阅读、生物节律、接触治疗、潜意识自助录音带和通灵侦探等）所宣称的事实，其实早已被证明是谬误的。本书提到，这些伪心理学产业的存在，助长了媒体的炒作式报道。这种趋势对心理学的危害远比对其他学科的危害要大，理解个中缘由，是学会正确看待心理学的一个重要环节。

本书面向的不只是即将成为心理学研究者的人，而是一个更大的读者群——**心理信息的消费者**。本书的目标读者是心理学初学者，以及那些在大众媒体上接触过一些心理学话题，又想知道如何去判断这些信息是否合理的广大读者们。

本书不是一本标准的心理学导论教材，它没有对心理学领域已取得的研究成果进行总结。事实上，仅仅到大学里选修一门心理学导论课程，对于纠正传媒对心理学的不实描述来说，可能并不是终极的解决方案。众多对心理学抱有很大兴趣的非专业人士，没有时间、没有钱或是没有机会进入大学进行正规的学习。更重要的原因是，作为一名大学心理学的教师，我不得不承认我和同事们也不总能引导初学者对心理学这门科学产生一个正确的认识。这是因为，初级的课程设置中通常没有包含批判性分析思维技巧的训练，而这正是本书讨论的焦点所在。作为教师，我们常常只会关注教学内容及课程内容的完备性。每当我们稍微偏离教学大纲，和学生讨论一些诸如媒体眼中的心理学话题时，都会感到有些许内疚，并开始担心自己会不会因为跑题而不能在学期结束前完成所有授课内容。

现在的心理学导论类教科书通常都有600~800页，并且引用了数百篇已经发表的文献。当然，包含如此丰富的材料并没什么错，它反映出心理学知识在不断增长。不幸的是，负面效果也同时存在。教师们常常只忙于给学生灌输一大堆的理论、事实和实验，而没时间去关注那些会

被学生带入心理学研究的基本问题及错误观念。这主要是因为教师们（包括导论类教科书的作者）想当然地认为，只要学生接触了足够多的心理学研究，自然就能从中推导出问题的答案。简言之，他们希望学生可以从对心理学各领域实证研究的讨论中，自行挖掘出各类问题的隐含答案。但是这类希望往往都落空了。到这门课的最后复习阶段或学期结束时，教师们才无比震惊和沮丧地发现，学生提出的一些问题及说法，是他们在课程开始第一天就应该提出来讨论的，而不是在 14 周之后。比如："心理学实验不同于现实生活，它们能告诉我们什么呢？"；"心理学无法像化学那样成为一门真正的科学，是吗？"；"可是，我在电视上听一位心理治疗师讲的正好与我们教科书上说的相反"；"我认为这个理论相当愚蠢——我弟弟的行为和这个理论所说的截然相反"；"心理学不过是些常识"；"每个人都知道什么是焦虑——何苦还要去定义它呢？"；"心理学不过是一堆观点而已"。对于很多学生来说，仅靠思考心理学教科书中的内容是无法为这些问题找到答案的。在本书中，我将对这类问题和说法背后的误区进行澄清。

不幸的是，研究发现，一般的心理学入门课程并不能有效地纠正初学者对心理学所持的诸多误解（Best，1982；Higbee & Clay，1998；McCutcheon，Furnham，& Davis，1993；Vaughan，1977）。曾有一位研究者说："我不得不说，那些'导论'性质的课程基本无法影响学生的错误观念"（Vaughan，1977，p. 140）。他还进一步得出结论说："目前我们还未能普遍地提高学生的批判性思维能力和意识，未能使学生有能力对没有证据的论断提出质疑"（p. 140）。沃恩后面的这个结论与本书撰写的初衷很吻合，心理学也许比其他科学更需要批判性的思维能力，这种能力使学生能够辨别良莠，将真正的科学与其他伪科学区别开来。这种批判性思维是学生必须掌握的，这样才能独立地评估心理学信息。

即使若干年后学生们不再记得心理学入门课程中的内容，他们仍然可以运用本书所涵盖的基本原理去判断心理学的言论。即使埃里克森（Erikson）的人生发展阶段论被忘得一干二净，他们也仍可以运用本书介绍的思维工具去辨别媒体中出现的心理学信息的真伪。一旦掌握，这

些技能可以成为终身受用的思维工具，帮助我们去评判各种知识主张。首先，它将使我们能够对某些观点的合理性做出一个初步和总体的判断。其次，这些技巧提供了一些评估“专家”观点可信度的标准。因为在复杂的社会中，人们对专家观点的依赖始终存在，在获取知识时，对专家观点可信度的判断就变得尤为重要了。虽然这些批判性思维技巧可运用于各个学科或知识门类，但它们在心理学领域里尤其重要，因为此领域经常被大众媒体所歪曲。

许多心理学家都对能否阻止心理学被歪曲持悲观态度。虽然这种悲观并非没有道理，但是这本类似“消费者指南”式的书源自于一个信念，那就是心理学家不能让这一问题成为一种自我实现式的预言。

尽管我很高兴这本《对“伪心理学”说不》能有多次再版的机会，但令人遗憾的是，本书存在的原因仍和本人当初撰写本书第一版时完全一样。媒体对心理学的介绍一如既往地在误导大众，而学生在开始上心理学入门课程时，还是带着与以往同样多的对于心理学的误解。正因为如此，本书后续几个版本的目标始终如前。这些目标正被越来越多心理学教师所共有。斯坦福大学心理学家罗杰·夏佩德（Roger Shepard）表达了与本书第一版写作初衷相同的看法：“虽然大多数心理学的本科生或许不会走上学术科研的道路，但我们仍然希望他们有能力去对那些不断出现在媒体上的片面、幼稚、混乱及夸张的所谓社会科学“发现”做出判断……那些广为流传、认为通过无根据的常识或更为糟糕的星相学之类的伪心理学就能充分理解人类行为和心理现象的谬论，向我们提出了持续的挑战”（Shepard，1983；p. 855）。

批判性思维技能帮助人们更好地理解心理学的主题，理解他们周围的世界所发生的事情。本书的目的即是对此技能作一个简略的介绍。

第8版更新的内容

第8版的《对“伪心理学”说不》在结构上没有作很大的改动，因

为之前的版本已经进行了章节重组。各章节的内容和顺序也保持原样。应评审者和读者的要求，这一版与第 7 版篇幅相同。读者和使用者都不希望本书篇幅增加，实际上，也确实没增加。第 8 版和第 7 版在篇幅上相同，比第 6 版缩减了 15% 左右，然而最好的例子（依据评审者和读者的反馈）都被保留了下来。最重要的是，我持续更新和修订了书中的许多例子，用最新的研究和论题替换掉了那些过时的事例。我最主要的努力在于引用与本书所提及的各种概念和实验结果相关的最新研究资料。结果就是，有 135 条新的引文出现在本书中，以便读者获得有关例子和概念的最新参考文献。

本书的出版宗旨一如既往——对批判性思维技能作简略介绍，帮助学生更好地去理解心理学的主题。在 20 世纪 90 年代，大学里强调批判性思维技能的呼声越来越高（Halpern，1998）。的确，一些州立大学系统已经进行了以加强批判性思维教育为目的的课程改革。与此同时，也有其他教育学学者认为，批判性思维技能不应该脱离特定的学科内容。而《对“伪心理学”说不》正好融合了这两种观点，帮助教师在教授丰富的现代心理学知识的同时传授批判性思维的技巧。

欢迎读者将自己对本书的意见寄到下面的地址：Keith E. Stanovich，Department of Human Development and Applied Psychology，University of Toronto，252 Bloor St. W.，Toronto，Ontario，Canada，M5S 1 V6. E-mail：KStanovich@oise.utoronto.ca。

致 谢

我在以前的版本中感谢过的许多人继续为这本书贡献着有益的想法。不过，我必须对詹姆斯·麦迪逊大学理查德·韦斯特（Richard West）单独致敬，他在这本书的改进过程中持续做出最有价值的贡献。他是一位具有人文精神的学者，一个真正的朋友，他在智力和情感上的支持令我感激万分。

还有一些学者为这一版本和早先的版本提供了有价值的反馈。他们包括：Wayne Bartz，American River College；Christopher Bauer，University of New Hampshire；Ludy Benjamin，Texas A&M University；Angela M. Birkhead-Flight，University of Cincinnati；Virginia Blankenship，University of Northern Arizona；Edward C. Chang，Northern Kentucky University；Michael Choban，West Virginia Wesleyan University；Jim Coan，University of Arizona；Ellen Cole，Alaska Pacific University；Ellen Cotter，Georgia Southwestern State University；Anne Cunningham，University of California，Berkeley；Ian Deary，University of Edinburgh；Julie Deisinger，Saint Xavier University；David DiBattista，Brock University；Wallace Dixon，Heidelberg College；Mark Fineman，Southern Connecticut State University；Herbert Fink，SUNY–Brockport；Heinz Fischer，Long Beach City College；Ronald Gandelman，Rutgers University；Michael Gasser，University of Northern Iowa；Traci A. Giuliano，Southwestern University；William Graziano，Purdue University；Nancy J. Gussett，Baldwin-Wallace College；Gordon Hammerle，Adrian College；Randy Hansen，Oakland University；George Heise，Indiana University；Albert Heldt，Grand Rapids Junior College；Dori Henderson，Metropolitan State University；George Howard，University of Notre Dame；Barry Kendall；Bernie Koenig，Fanshawe College；Victor Koop，Goshen College；P. A. Lamal，University of North Carolina，Charlotte；Stephen Louisell，Kalamazoo Community College；Margaret Matlin，SUNY-Geneseo；Douglas Mook，University of Virginia；Timothy Moore，York University；Edward Morris，University of Kansas；Joseph E. Morrow，California State University at Sacramento；Michael O' Boyle，Iowa State University；Blaine Peden，University of Wisconsin，Eau Claire；John F. Pfister，Dartmouth College；Sam Rakover，University of Haifa；Richard Redding，Hahneman University；Michael Ross，University of Waterloo；John Ruscio，Elizabethtown

College；Walter Sa，Grand Valley State University；Allen Salo，University of Maine at Presque Isle；Frank Schieber，University of South Dakota；Marjorie Semonick，University of Minnesota；David Share，University of Haifa；Jeffrey Sherman，Northwestern University；Linda Siegel，University of British Columbia；Norman Silverman，University of Illinois，Chicago；Frank Smoll，University of Washington；Paul Solomon，Williams College；Mike Stadler，University of Missouri；Maggie Toplak，York University；Larry Vandervert，Spokane Falls Community College；John Vokey，University of Lethbridge；Carol Wade，College of Marin；Marty Wall，University of Toronto；Barbara Wanchisen，Baldwin-Wallace College；Toni G. Wegner，University of Virginia；Edward Wisniewski，Northwestern University；Murray S. Work，California State University at Sacramento；and Edward Zuckerman，Guilford Press。书中许多灵感都来自于与奥克兰大学的 Ted Landau，Larry Lilliston 和 Dean Purcell 关于教学方法学的讨论。

感谢 Robyn Macpherson 在图书资料和参考文献方面的大力协助，使得第 8 版能够按时付梓。Carol Kelley 鼓舞了我写作的士气。感谢 Andrée Debruge 最后时刻的帮助。我要感谢以下为这个版本提供了宝贵建议的评阅者：Steven Isonio，Golden West College；Cindy Sifonis，Oakland University；Michael Tagler，Nebraska Wesleyan University；and Chris Ward，Stonehill College。Allyn & Bacon 出版社的编辑 Susan Hartman 为本书倾注了指导、热情和支持，她的编辑助理 Therese Felser 也是如此。

最后，我还要感谢 Paula J. Stanovich，不仅仅是出于所有致谢辞中都会提到的情感支持。她对全人类，特别是那些不幸者的关注，激励着所有认识她的人。我们都认同一个理念：所有的人都应该有机会充分发挥自己的潜力。这本书，足以证明我有这样的机会。这种机会终将泽被包括残障人士在内的所有人，而 Paula 的工作正在让这一天能够早日到来。

简要目录

序　言　xxi

第1章　心理学充满生机（在科学阵营里左右逢源）　1

弗洛伊德的理论只是心理学中极小的一部分，已经远远不能代表作为科学的心理学。

第2章　可证伪性——如何捕捉头脑中的小精灵　**22**

如果一个理论不可证伪，并且与自然界的真实事件没有关联，那么它就是无用的。

第3章　操作主义和本质主义——“但是，博士，这到底是什么意思？”　**40**

心理学并不试图回答本质主义的终极问题，而是将概念建立在可观测的现象基础之上。

第4章　见证和个案研究证据——安慰剂效应和了不起的兰迪　**60**

伪心理学所依赖的个案研究和见证叙述为何是毫无价值的？

第5章　相关和因果——用“烤箱法”避孕　**83**

两个变量之间存在相关，并不能保证一个变量的变化就会导致另一个的变化。

第6章　让一切置于控制之下——聪明汉斯的故事　96

心理学家通过创设特殊条件，对实验进行操纵和控制，保证结论的科学性。

第7章　“但是这不是真实的生活！”——“人为性”批评和心理学　117

心理学实验常常被批评不能代表“真实的生活”，这种批评为何是不合理的？

第8章　避免爱因斯坦综合症——聚合性证据的重要性　134

心理学的进步靠的不是一两次突破，而是不断积累、整合而渐进发展的。

第9章　打破“神奇子弹”的神话——多重原因的问题　157

大部分复杂的行为都是多重原因决定的，所谓的“神奇子弹”并不存在。

第10章　人类认知的阿喀琉斯之踵——概率推理　166

心理学中的结论是概率式的——大部分情况下会发生，但是并非所有情况下都发生。

第11章　偶然性在心理学中扮演的角色　184

某些巧合的发生只是纯属偶然，并不需要特别的解释。

第12章　不招人待见的心理学　205

伪心理学在大众媒体和图书中日益泛滥，如何识别伪心理学，学会对伪心理学说不？

参考文献　230

目　录

序　言　xxi

第1章　心理学充满生机（在科学阵营里左右逢源）　1

弗洛伊德问题　1
现代心理学的多样性　3
　多样性的含义　5
科学的统一性　7
那么，什么是科学?　10
　系统的实证主义　10
　公共性的、可检验的知识：可重复性和同行评审　12
　实证性的、可解决的问题：科学家对于可检验理论的研究　14
心理学和世俗智慧："常识"的误区　16
心理学是一门年轻的科学　19
小结　21

第2章　可证伪性——如何捕捉头脑中的小精灵　22

理论和可证伪性标准　23
　敲门节奏理论　25
　弗洛伊德与可证伪性　26
　小精灵　28

不是所有的证实都等价　30
可证伪性和世俗智慧　31
承认错误的自由　32
想法不值钱　35
科学中的错误：接近真理　36
小　结　39

第3章　操作主义和本质主义——“但是，博士，这到底是什么意思？”　40

为什么科学家不是本质主义者　40
本质主义者喜欢咬文嚼字　41
操作主义者将概念和可观测事件联系在一起　42
信度和效度　44
直接和间接的操作性定义　46
科学概念的演进　46
心理学领域的操作性定义　48
作为人性化力量的操作主义　51
本质主义问题和对心理学的误解　54
操作主义和心理学问题的措辞　55
小　结　59

第4章　见证和个案研究证据——安慰剂效应和了不起的兰迪　60

个案研究的地位　62
为什么见证叙述毫无价值：安慰剂效应　64
“鲜活性”问题　68
单一个案的压倒性影响　72
了不起的兰迪：以彼之道，还施彼身　74
见证为伪科学打开方便之门　77

小 结 81

第5章 相关和因果——用“烤箱法”避孕 **83**

第三变量问题：戈德伯格与糙皮病 84
为什么戈德伯格的证据更好 85
方向性问题 89
选择性偏差 90
小 结 95

第6章 让一切置于控制之下——聪明汉斯的故事 **96**

斯诺与霍乱 97
比较、控制和操纵 98
随机分配与操纵共同定义了真实验 100
控制组的重要性 102
聪明汉斯——神马的故事 104
20世纪90年代的聪明汉斯 106
对变量分开考察：特殊条件 110
直觉物理学 112
直觉心理学 114
小 结 115

第7章 “但是这不是真实的生活！”——“人为性”批评和心理学 **117**

为什么自然性并非总是必要的 117
对“随机取样”的误解 119
随机分配和随机样本的区别 120
理论研究和应用研究的异同 121
心理学理论的应用 125
“大二学生”问题 127

正确看待“真实生活”和“大二学生”问题 132
小 结 133

第 8 章 避免爱因斯坦综合症——聚合性证据的重要性 134

关联性原则 135
消费者规则：警惕关联性原则的无效性 136
“跃进”模式与渐进整合模式的比较 139
聚合性证据：在瑕疵中进步 140
心理学中的聚合性证据 142
科学共识 146
研究方法和聚合性原则 148
向更有效的研究方法迈进 150
不要对矛盾数据感到绝望 152
小 结 156

第 9 章 打破“神奇子弹”的神话——多重原因的问题 157

交互作用 159
单一原因解释的诱惑 161
小 结 165

第 10 章 人类认知的阿喀琉斯之踵——概率推理 166

“某某人”统计学 168
概率推理以及对心理学的误解 169
有关概率推理的心理学研究 171
对概率信息的不充分利用 172
样本大小信息的误用 174
赌徒谬误 176
再谈统计与概率 181
小 结 182

第 11 章 偶然性在心理学中扮演的角色 184

试图解释偶然性事件的倾向 184

解释偶然性：错觉相关和控制错觉 188

偶然性和心理学 191

巧 合 191

个人的巧合 195

接受错误以减少错误：临床预测与统计预测 196

小 结 203

第 12 章 不招人待见的心理学 205

心理学的形象问题 205

心理学和超自然心理学 206

自助类读物 209

菜谱式知识 210

心理学和其他学科 212

我们是自己最坏的敌人 214

每个人不都是心理学家吗 221

抵制科学心理学的根本原因 223

结束语 228

参考文献 230

1

心理学充满生机
（在科学阵营里左右逢源）

弗洛伊德问题

在街上随便拦住 100 个人，让他们说出一个活着的或已故的心理学家的名字，然后记下他们的答案。毫无疑问，他们提到的会是菲尔博士（Dr. Phil）、韦恩·戴尔（Wayne Dyer）及其他一些“媒体心理学家”。如果我们把这类媒体和通俗心理学家排除在外，只考虑那些对心理学做出过卓著贡献的心理学家，那么这项非正式小调查的结果就几乎没什么悬念了——西格蒙德·弗洛伊德（Sigmund Freud）会名列榜首，B. F. 斯金纳（B. F. Skinner）可能会屈居次席，但远落后于弗洛伊德。没有任何其他的心理学家具有足够的知名度来撼动这两位的地位。因此可以说，弗洛伊德和那些在媒体上频频露脸的通俗心理学共同定义了公众心目中的心理学。

弗洛伊德的声名远播，极大地影响了普通公众心目中关于心理学的概念，同时也造就了诸多认识上的误区。例如，许多刚入门的心理学学生会惊讶地发现，如果对 APA（American Psychological Association，美国心理学会）会员中所有认同弗洛伊德精神分析的人数进行一下统计，

他们的人数居然没有占到会员总数的 10%。在另一个主要的心理学组织 APS（Association for Psychological Science，美国心理协会）中，这一比例也不会高于 5%。

现代心理学并没有像媒体和一些人文学科那样被西格蒙德·弗洛伊德的理论所左右，也没有被其所限定。在现代心理学家所关注的大量研究主题、数据和理论中，弗洛伊德的工作只占其中极小的一部分，在这些研究和理论中占更大比重的则是近期 5 位诺贝尔奖得主所做的工作：大卫·胡贝尔（David Hubel）、丹尼尔·卡尼曼（Daniel Kahneman）、赫伯特·西蒙（Herbert Simon）、罗杰·斯佩里（Roger Sperry）和托斯腾·维瑟（Torsten Wiesel）以及美国国家科学基金的前负责人理查德·阿特金森（Richard Atkinson）的贡献，然而，这些人的名字对公众来说却是颇为陌生的。

弗洛伊德对于现代心理学的重要性被无限地夸大了，这就已经足够糟糕了。更糟的是，弗洛伊德的调查方法完全不能代表现代心理学家是如何进行研究的（弗洛伊德开展其著名工作已经是一百多年前的事儿了）。事实上，弗洛伊德式的研究方法彻底误导了人们对心理学研究的印象。例如，弗洛伊德并不采用控制实验，而我们将在第 6 章讲到，控制实验是现代心理学家“兵器库”中最有力的武器。弗洛伊德认为，个案研究足以证明理论的真实或谬误，在第 4 章中，我们将谈谈这一理念为何是错误的。最后，弗洛伊德的工作中最大的问题是理论和研究数据的联系。正如我们将在第 2 章看到的，对于一个科学理论来说，理论和研究数据的联系必须满足一些标准，而弗洛伊德的理论常常不能满足这些标准（Crews，1996，1998；Hines，2003；Macmillan，1997；McCullough，2001；Watters & Ofshe，1999）。简而言之，弗洛伊德根据他得到的数据（个案研究和内省）建立了一套精细的理论，而这些数据并不足以支撑此理论；他专注于构建复杂的理论构架，但并没有像许多现代心理学家那样，保证这些理论建立在可靠、可重复的因果关系之上。总之，人们对于弗洛伊德式的工作太过熟悉，这严重阻碍了他们对现代心理学的正确理解。

在这一章中，我们将采用两种方法来解决“弗洛伊德问题”。首先，当我们展示现代心理学的多样性时，就能非常清楚地了解到弗洛伊德的工作所占的比重其实是很小的（见 Haggbloom et al.，2002；Robins，Gosling，& Craik，1999，2000）。其次，我们将讨论一下，广泛而多样的心理学研究中，哪种特征是最为普遍的。有关弗洛伊德工作的那种过时的认识遮蔽了普通大众的双眼，使之无法看到现代心理学所共有的唯一而普遍的特性：用科学的方法寻求对行为的理解。

现代心理学的多样性

事实上，现代心理学包含了大量不同的内容和观点。这种多样性使得作为一门学科的心理学显得不那么浑然一体。美国心理学基金会杰出教学奖得主亨利·格雷特曼（Henry Gleitman，1981）将心理学描述为“一个松散地联合在一起的学术王国，它横跨了生物科学和社会科学两个领域”（p. 774）。

心理学有着令人惊叹的广泛性和多样的调查方法，知道这一点对于理解心理学的本质至关重要。可以简单列举一些具体指标来证明这一点：美国心理学会（APA）有 53 个分支机构，每个分支都代表了一个特定的研究或应用领域（见表 1.1）。从表中你可以看到心理学研究主题、研究背景和研究方法的丰富性和多样性。另一个大型心理学组织——美国心理协会（APS）也同样分支众多。其实，表 1.1 对于心理学领域的多样性的描述还是较为保守的，因为它给我们造成了一种印象，即每个分支都是一个特定的专业领域。事实上，53 个分支机构中的每一个都是非常宽泛的研究领域，包含更小的不计其数的分支！简言之，要穷尽心理学领域主题的多样性是非常困难的。

表 1.1 美国心理学会（APA）的分支机构

1. 普通心理学（General Psychology）
2. 心理学教学（Teaching of Psychology）
3. 实验心理学（Experimental Psychology）
5. 评价、测量和统计（Evaluation，Measurement, and Statistics）
6. 行为神经科学和比较心理学（Behavioral Neuroscience and Comparative Psychology）
7. 发展心理学（Developmental Psychology）
8. 人格和社会心理学（Personality and Social Psychology）
9. 社会问题的心理学研究（Psychological Study of Social Issues）
10. 审美、创造力及艺术心理学（Psychology of Aesthetics，Creativity，and the Arts）
12. 临床心理学（Clinical Psychology）
13. 应用咨询心理学（Consulting Psychology）
14. 工业和组织心理学（Industrial and Organizational Psychology）
15. 教育心理学（Educational Psychology）
16. 学校心理学（School Psychology）
17. 理论咨询心理学（Counseling Psychology）
18. 公共服务中的心理学家（Psychologists in Public Service）
19. 军事心理学（Military Psychology）
20. 成人发展与老龄化（Adult Development and Aging）
21. 应用实验和工程心理学（Applied Experimental and Engineering Psychology）
22. 康复心理学（Rehabilitation Psychology）
23. 消费者心理学（Consumer Psychology）
24. 理论和哲学心理学（Theoretical and Philosophical Psychology）
25. 行为分析（Behavior Analysis）
26. 心理学史（History of Psychology）
27. 社区心理学（Community Psychology）
28. 精神药理学和药物滥用（Psychopharmacology and Substance Abuse）
29. 心理治疗（Psychotherapy）
30. 心理催眠（Psychological Hypnosis）
31. 国家心理学会事务（State Psychological Association Affairs）
32. 人本主义心理学（Humanistic Psychology）
33. 智力缺陷和发展性障碍（Mental Retardation and Developmental Disabilities）
34. 人口与环境心理学（Population and Environmental Psychology）
35. 女性心理学（Psychology of Women）
36. 宗教心理学（Psychology of Religion）
37. 儿童、青少年和家庭服务（Child，Youth，and Family Services）

表 1.1 美国心理学会（APA）的分支机构（续表）

38. 健康心理学（Health Psychology）
39. 心理分析（Psychoanalysis）
40. 临床神经心理学（Clinical Neuropsychology）
41. 心理学和法律（Psychology and Law）
42. 独立从业的心理学者（Psychologists in Independent Practice）
43. 家庭心理学（Family Psychology）
44. 男女同性恋及双性恋的心理学研究（Psychological Study of Lesbian, Gay, and Bisexual Issues）
45. 少数民族的心理学研究（Psychological Study of Ethnic Minority Issues）
46. 媒体心理学（Media Psychology）
47. 锻炼和运动心理学（Exercise and Sport Psychology）
48. 和平心理学（Peace Psychology）
49. 团体心理学和团体治疗（Group Psychology and Group Psychotherapy）
50. 成瘾（Addictions）
51. 男性和男性化的心理学研究（Psychological Study of Men and Masculinity）
52. 国际心理学（International Psychology）
53. 临床儿童心理学和青少年心理学（Clinical Child Psychology and Adolescent Psychology）
54. 幼儿心理学（Pediatric Psychology）
55. 药物疗法（Pharmacotherapy）

注：没有分支 4 和 11。

多样性的含义

许多人学习心理学是希望能够学到一套宏大的心理学理论，以此来概括和解释人类行为和意识的方方面面。但这类愿望总是会落空，因为构成心理学的不是一整套宏大的理论，而是许许多多不同的理论，每个理论仅仅能够解释行为的有限方面（Benjamin, 2001；Griggs, Proctor, & Bujak-Johnson, 2002；Zechmeister & Zechmeister, 2002）。心理学的多样性使得理论整合变得极为困难。事实上，在许多心理学者看来，“整合”本身就是“不可能的任务”。尽管如此，另外一些心理学家却正

在寻求领域内的理论整合（Henriques，2003，2004；Kenrick，2001；Kimble，1999；Sternberg，2005）。例如，在过去的10年间，心理学的学科统一性有所增强，这要归功于进化心理学家的努力。这些研究者将人类心理过程视为服务于某些重要进化功能（诸如亲缘关系识别、伴侣选择、合作、社会交换及后代抚养等）的机制（Barrett，Dunbar，& Lycett，2002；Bjorklund & Pellegrini，2002；Buss，2003；Cosmides & Tooby，2000；Geary，2005；Pinker，2002），并试图以此来实现概念的整合。

然而，无论心理学家对心理学主题的统一性持何立场，他们都承认，即便有一天能够实现理论的整合，其过程也是极为困难的。缺乏理论的整合为心理学招来了一些批评，贬低了它作为科学取得的进步。这类批评源于一个错误的观念，即所有真正的科学都必须具备一个宏大的、统一的理论。之所以说它错误，是因为它忽视了其他科学同样也缺乏一个完备统一的概念体系这个事实。哈佛大学心理学家威廉·艾斯特斯（William Estes，1979）已经强调过这一点：

> 实验心理学家所面临的这种困境既不新鲜，也非独有。20世纪早期，物理学在本科水平的教学中便被分成了若干独立学科。因此，我是通过分别学习力学、热力学、光学、声学和电学而了解这门科学的。同样，化学也曾被分为无机化学、有机化学、物理化学和生物化学。当时这些分支之间的交流和融合并不比现在的心理学好到哪里。它们都仅仅在抽象的数学理论水平上才实现了整合。医学也被分为众多分支，而且和心理学一样，没有新的整合出现。（pp. 661–622）

事实上，心理学分化得如此严重，以至你不难想象，如果一所大学撤掉心理系，它会很容易将心理学系的成员分派到其他院系去。生理心理学家可以去生物系；社会心理学家可以去社会学系；认知心理学家可以去认知科学的交叉学科院系；工业与组织心理学家可以去商学院；临床和咨询心理学家可以去社会工作、人力资源和教育系；发展心理学家

可以去教育、认知科学或人力资源系等等。这些心理学家在与新同事进行学术交流时几乎不会发觉有什么隔阂。实际上，许多心理学家反而觉得新同事更易于共事。从内容上来说，现代心理学绝不仅仅是多个主题的简单统一体，人们必须在更高的层次上寻找整合这一学科的因素。

一旦我们知晓了决定某一门学科结构的社会和历史因素是怎么回事，就能认识到，要求所有领域具备高度统一性是不合逻辑的。事实上，“心理学研究”这一术语要比“心理学”一词更能准确反映这一学科的多样性。而且用“心理学研究”能让学生不再惊讶于同一个学科的不同领域取得科学进步的速度竟然如此迥异—— 一些领域在解释和预测行为方面取得了显著的成就，而另外一些则成果寥寥。“心理学”这个词显然不能承载和传达这一差异化的事实，相反，它所表达出的“统一性”内涵恰恰是这门学科所不具备的。

如果我们试图找到心理学研究课题间的统一性，那就不要继续在心理学研究的内容之间找寻所谓的联系，我们应当关注心理学家获取新知识所采用的方法。这是我们唯一有希望在心理学家之间找到共性的地方。但即使是在方法领域，也依然存在一些有关这门学科的很深的误解。

科学的统一性

仅说心理学是关于人类行为的科学，并不能将它和其他学科区分开来。许多其他专业团体和学科——包括经济学家、小说家、法律、社会学、历史、政治科学、人类学和文学研究——都或多或少与人类行为有关，心理学在这方面并非独树一帜。

应用性也不能证明心理学具有任何独特性。例如，许多大学生选择主修心理学是因为他们有一个要帮助他人的崇高目标。但是在许多领域，如社会工作、教育、护理、职业咨询、物理治疗、警事科学、人力资源以及语言矫正等，“帮助他人”都是其中的重要组成部分。同样，“通过提供咨询来帮助他人”也是众多其他领域的重要组成部分，这些领域包

括教育、社会工作、警事工作、护理、神职工作、职业咨询等等。培训应用性的、通过咨询来帮助他人的专业人才并不需要单独开辟一门叫做“心理学”的学科。

只有两点能证明心理学是一门独立的学科。其一，心理学研究采用科学方法来探究人类及动物的所有行为；其二，从这一知识衍生出的实际应用是具备科学基础的。如果不是这样，心理学就失去存在的理由了。

心理学不同于其他行为研究领域的地方在于，它试图向公众保证两点：第一，心理学中有关行为的结论都有科学证据；第二，心理学的应用都源于科学方法，并经过了科学方法的检验。心理学是否曾经偏离过这两个目标呢？有过，而且经常如此（Lilienfeld，Lynn，& Lohr，2003；Loftus & Guyer，2002；Lynn，Loftus，Lilienfeld，& Lock，2003；Mook，2001；Watters & Ofshe，1999）。本书就是关于怎样更好地实现这两个目标的。在第 12 章中我将回到这一主题——一些心理学工作者因为不遵守适当的科学标准而自我损害了其作为心理学家的合法性。但是，从原则上讲，科学性正是保证心理学作为一门独立学科的标准。如果有朝一日心理学不再追求这些目标——即它不再愿意坚守科学标准——那它也就应该关张大吉，将其关注的领域拱手让于先前提到的那些其他学科——因为此时它已成为了一个完全多余的知识领域。

无疑，任何人想要理解心理学，第一步，也是至关重要的一步，就是要意识到心理学的首要特征——它是有关行为的、以数据为基础的科学研究。对这一事实及其全部内涵的理解将贯穿本书，因为这是我们认识真正心理学的最基本的途径。反过来说，人们之所以对心理学的理解会出现各种各样的偏差，正是因为未能意识到它是一门科学的学问。例如，我们常常会听到学术圈外的人宣称心理学不是科学。为什么还会有这样的误解？

那些想让公众相信“心理学不能成为一门科学”的企图，其产生的背景各不相同。正如我们在后面的章节中所要讨论的，许多有关心理学的错误认识，都是由那些伪心理学的代理人处心积虑制造的。在我们的社会中，一个经营伪科学信念系统的巨大产业正在兴起，这一信念系统

出于既得利益的考虑，总是想让大众相信，无论什么都能纳入心理学的范畴，而且心理学的主张不能以理性标准来衡量。这无疑为"催眠减肥"、"激发潜在心灵能量"、"睡觉时学法语"这类广告以及利润高达数百万的"心理自助"产业里其他诸多门道的营销创造了绝佳的氛围。此类门道要么不是建立在科学证据基础上，要么（在许多时候）与已有的证据相冲突。

另一种对于科学心理学的排斥是由于，一些人不愿看到科学进入到长期以来由不容置疑的权威或"常识"统治的领域里。历史上此类例子不胜枚举——人们拒绝使用科学，而更喜欢利用哲学沉思、神学谕告或世俗智慧去解释现实世界。每一门科学都会经历一个受到阻碍的阶段。与伽利略同时代的知识分子拒绝透过他的新望远镜观察天空，因为"木星存在卫星"颠覆了他们的哲学和神学信仰。几个世纪以来，人类解剖学的发展可谓步履蹒跚，因为世俗和宗教禁止对人类尸体进行解剖（基督徒认为，身体的内部是"上帝的辖区"，见 Grice，2001）。查尔斯·达尔文总是被反复抨击。保罗·布洛卡（Paul Broca）的人类学协会（Society of Anthropology）在法国受到抵制，因为有人认为，关于人类的知识会颠覆国家。

关于人类的知识每向前迈进一步，都会引发反抗。然而，当人们开始意识到科学并没有通过调查和研究对人性造成亵渎，而是以扩展知识的方式促进了人类的自我实现时，反抗终将烟消云散。谁现在还认为星系图以及宇宙是由无数星球所组成的复杂理论会摧毁我们对于宇宙的向往？谁会选择禁止人体解剖时的医疗保健系统，并进而拒绝从社区中获得的现代医疗保健呢？对于星球和人类身体的实证性态度并没有磨灭人性。更近的例子是，达尔文的进化论体系为遗传学和生物学取得非凡的进步奠定了基础。但是，在我们更接近人类的本质及起源的同时，残余的反抗势力仍然存在。在美国，宗教鼓吹者继续施压，意欲在公立学校推行"神创论"教学；同时，调查显示，有很大比例的民众（在某些调查中，是大多数）并不接受"人类是经过自然选择进化而来的"这一科学事实（Lerner，2005）。进化生物学有着无数辉煌的科学成就记录，

时至今日还是照样被公众所排斥。如此看来，心理学——这门志在将所有关于人类的固有信念都置于科学检验之下的新兴科学，时下还会引发人们对其正确性的否定，这又有什么好奇怪的呢？

那么，什么是科学？

为了理解什么是心理学，我们必须理解什么是科学。或许我们可以从“什么不是科学”入手。按这种方法，我们能摒弃大部分常见的错误观念。首先，科学并不是由内容来定义的。宇宙万物的任何方面对于一门科学学科的发展来说，都是一场公平的游戏，当然也包括人类行为的所有方面。我们不能将宇宙万物分为“科学的”和“非科学的”两类。尽管历史上始终有一股强大的力量，试图将人类排除在科学研究的范围之外，但正如我们所见，它们均以失败告终。拒绝将心理学作为一门科学学科来对待，可能代表了这一历史争论的余音。

科学也不能按照特定实验器材的使用来定义。试管、电脑、电子设备或研究者的白大褂都定义不了科学（即使这些成为衡量标准，心理学的科学地位也是无懈可击的，因为所有大学的心理系都充斥着电脑、药剂和各种型号的电子设备）。这些都是科学的附属物而不是其本质特征。科学，是一种思考和观察事物以便深入理解其运行机制的方法。

在本章的剩余部分，我们将讨论科学的三个相互关联的重要特征：（1）应用系统的实证主义；（2）产生公共知识；（3）验证可解决的问题。尽管我们将逐一检验每一条特征，但请记住这三条特征构成了相互联系的统一整体。（更多有关科学的普遍特征的详细讨论，参见书后参考文献部分列出的 Bronowski、Cournaud、Medawar、Popper、Raymo 以及 Sagan 的著作。）

系统的实证主义

如果在任何辞典中查找“实证主义”，你会发现它的意思是“基于

观察的实践”。科学家通过验证来找寻世界的规律。这个事实可能对你来说是显而易见的事实，而这正是过去两个世纪以来科学态度传播的结果。在过去，它却不是显而易见的。回想一下不愿透过伽利略的望远镜看星空的人。长久以来，人们认为获得知识的最佳途径是纯粹思考或诉诸权威。伽利略宣称看见了环绕木星的卫星。另一位学者，弗朗西丝科・西奇（Francesco Sizi），试图驳倒伽利略，但不是通过观察，而是通过下面的一番话：

> 人的脑袋上有七个窗口：两个鼻孔，两只耳朵，两只眼睛和一张嘴；因此在天界有两颗吉星，两颗灾星，两颗发光星（指日月——译者注），以及性状不明但无关紧要的水星。从这点和其他无数相似的自然现象诸如七种金属等等中，我们就可以归纳出行星必然是七个……除此之外，犹太人和其他古老的民族，都将一周分为七天，并以七大行星来命名；如果现在我们增加了行星的数目，将导致整个系统的崩溃……进一步来说，卫星用肉眼无法看到，因此对于地球没有影响，既然没有用处，也就不存在。（Holton & Roller, 1958，p. 160）

关键问题不在于以上论述多么愚蠢可笑，而在于它被视做可与真实观察抗衡的一种辩驳！今天我们嘲笑它是因为我们都是事后诸葛亮。三个世纪以来，业已证明力量的实证取向使我们强于可怜的西奇。要是没有经历这些实证主义的岁月，我们中的许多人可能都会点头同意并对他大加褒奖。的确，实证取向并不一定显而易见，这就是为何即使在一个科学占统治地位的社会中，我们也不得不经常强调它的原因。

但纯粹、单一的实证主义还不够。注意本章的标题是“系统的实证主义”。观察很好，而且很有必要，但是对于自然世界单纯的、非结构化的观察并不能导致科学知识的产生。假使你记录下自己一天中从起床到睡觉之间观察到的所有情况，完成这一任务时，你会拥有一大堆事实，但仅此并不能让你更好地理解这个世界。科学观察被称为“系统性的”，是因为它是结构化的，所观察的结果能够揭示自然世界一些潜在的本质。

科学观察通常都是理论驱动的；它们检验有关世界的各种不同解释观点。它们是结构化的，因此可以根据观察结果，决定哪些理论得到支持而哪些则被拒绝。

公共性的、可检验的知识：可重复性和同行评审

从某种特殊意义上说，科学知识是公共性的。当然，并不是说把科学发现张贴在社区的公告板上就叫“公共性”了。我们指的是这样一个事实，即科学知识并不单独存在于特定个体的头脑之中。从某种重要意义上说，科学知识在没有提交给科学团体、接受他人批评和验证之前是根本不存在的。那些被认为隶属于特定个体思维过程中、不可接受他人审查和批评的“特殊”知识，永远都无法获得科学知识的地位。

科学通过可重复性来实现其公共可检验性的理念。一项发现如果想在科学界获得公认，就必须以一种能够让其他科学家尝试相同实验并获得相同结果的方式呈现给科学团体。当这一切都完成，我们就可以说，这一发现是可重复的。科学家利用可重复性来定义公共知识。可重复性保证了特定发现并不是由于个别调查者的错误或偏差而产生的。简而言之，一项发现如果想要被科学团体所接受，它必须能够被原始调查者以外的其他人所重复。当一项发现经历了这种方式呈现出来时，它就成为了公共性的。它不再仅仅为原始研究者个人所有，它能够被其他人获取、扩展、批评，或以他们的方式得到应用。

诗人约翰·唐尼（John Donne）告诉我们“任何人都不是一座孤岛”。在科学中，没有一个研究者是一座孤岛。每个研究者都与科学团体及其知识基础相联系。正是这种相互联系使得科学累积性地发展。研究者不断在原有知识的基础上进行新的探索，力求超越已知。而这一过程的前提便是，先前的知识以一种适当的方式予以陈述，使任何研究者都能以之为基础来进行探索。

公共性的、可检验的知识，指的是我们可以将研究发现递交给科学

团体，团体中的任何人都能对其进行重复检验、批评或拓展。这个标准不仅对于科学家，同时对于作为消费者的外行人来说也是最重要的，因为他们必须对来自媒体的科学信息进行评估。正如我们将在第 12 章所看到的，区分大搞伪科学的江湖术士和真正的科学家的最主要的一个方法就是，前者常常避开科学出版的常规渠道，而选择直接通过媒体公开他们的“发现”。当公众面对真实性可疑的科学发现时，一个屡试不爽、颠扑不破的标准就是，考察这些发现是否在得到认可的科学期刊上发表过，以及是否经过了同行评审？对这一问题的回答往往能够区分“李鬼”和“李逵”。

同行评审是指每一篇投到期刊的文章都要经过数位科学家的评审，并将批评意见提交给编辑（通常都是此期刊所覆盖的某一领域中有资深研究经历的专家），再由编辑权衡这些意见，确定这篇文章可以立即发表，还是需进一步实验研究和统计分析之后再发表，或是因为有缺陷或价值太低而拒绝接受。大多数期刊在每期中都刊有编辑条例说明，因此很容易知道此期刊是否经过同行评审。

并非所有经过同行评审的科学期刊中的信息都必然正确，但至少它已满足了同行批评和监督的标准。这只是一个最低标准，而非严格的标准，因为大多数学科领域中都会有几十种质量参差不齐的期刊。大部分科学观点在满足一些基本标准的前提下，都可以在某些地方以正规的方式出版。那种认为只有很小部分的数据和理论才能够在科学界获得出版的观点是错误的。当一些心理救助或治疗方面的江湖术士试图说服媒体和公众时，往往会暗示，一种所谓“正统科学”的阴谋将他们排除在科学出版渠道之外。但是，稍微想一想，心理学领域中有多少这样的合法渠道啊。出版物《心理学摘要》（*Psychological Abstracts*）罗列了来自 1000 多种不同期刊的文章摘要。其中大部分期刊都设有同行评审。几乎所有待检验的理论和实验都能在如此众多的出版物中找到自己的发表渠道。

再次强调，我不认为《心理学摘要》中引用的所有期刊上的所有观点都必然正确。相反，正如我先前所强调的，发表只是一个最低的标准。

然而关键是，任何一种理念、一个理论、一项主张或疗法如果不能在有同行评审的学科文献中获得适当的收录，问题就很明显了。尤其是当某一主张缺乏证据却伴随着媒体的宣传运动时，此理念、理论或疗法显然是骗人的。例如，2005 年宾夕法尼亚州有一桩著名的诉讼，有人试图在学校的生物课上教授神创论，鼓吹智力设计（神创论的一种形式）理念的一个证人说，他很难举出任何一个经过同行评审的有关智力设计的研究，尽管这一运动已经兴起了十年有余（Talbot，2005，p. 68）。

同行评审机制在不同学科之间有所区别，但是根本理念是相同的。同行评审是科学将客观性和公开评议标准制度化的一种方法（另一种是重复验证）。观点和实验要提交给其他评估者，经过一个仔细推敲的过程。只有通过这一严格过程的观点才算符合了公共验证的标准。同行评审过程绝非完美，但它对我们消费者来说是唯一的保护机制。忽视或轻视它，就等于让我们自己被巨大的伪科学产业玩弄于股掌之间，而这一产业又极其善于操纵媒体来达到自己的目的（见第 12 章）。在随后的章节中，我们将更详尽地讨论，如果忽视这些心理科学实践中固有的审查与制衡，我们将要付出多么高昂的代价。

实证性的、可解决的问题：科学家对于可检验理论的研究

科学针对的是可解决的、可明确具体的问题。这意味着就其类型来说，科学家们所致力解决的问题是能通过现有的经验技术获得答案的。如果在当前所掌握的经验技术条件下，问题无解或理论不可验证，则科学家们将不会对它展开研究。例如，“在日托期间给予结构化语言刺激的 3 岁儿童，与那些没有给予这些额外刺激的儿童相比，是否可以更早地做好接受阅读指导的准备？”就是一个科学问题，因为在现有的经验方法之下，这是一个可解的问题。“人性本善还是本恶？”就不是一个可实证的问题，因此不属于科学领域。“生命的意义是什么？”同样也不是实证问题，因此也不属于科学领域。

科学通过以下方式得以进步：提出理论解释世界中的特定现象，根

据这些理论做出预测，实证地检验这些假设，基于检验的结果对理论进行修正（通常次序为：理论 → 预测 → 检验 → 修正）。因此对于科学家来说，“可解问题”这个词的意义通常是“可检验的理论”。什么样的理论才算是“可检验”的呢？这一理论必须与真实世界中可观察的事件具有特定的关联；这就是“实证可检验”的含义。可检验性标准在学术上通常被叫做证伪标准，这也是本书第 2 章的主题。

我们说科学家解决实证可解问题，并不是说在不同类别的问题中，有的本质上是可以解决的，而有的则注定无法解决，并且这种区分是固定不变的。恰恰相反，有些当前无法解决的问题，在理论和经验技术更加进步的时候会成为可解的。例如，20 年前，对于“托马斯·杰斐逊是否与其奴隶萨丽·海明斯生下了她的某个孩子”这一争议话题，不会有历史学家认为它是一个实证可解的问题。然而到了 1998 年，由于基因技术的进步，这个问题已成为可解的，发表在《自然》（*Nature*）杂志的一篇文章（Foster et al.，1998）指出，杰斐逊极可能是埃斯顿·海明斯·杰斐逊的父亲。

这就是科学得以发展而新的科学得以诞生的方式。但对于“当前什么是可解的”这一问题，总是存在巨大分歧。因为涉及正处于模棱两可状态的问题，科学家们自己在这点上的意见都难以统一。因此，尽管所有科学家都认同可解性标准，但是他们对其特定应用可能存在不同的意见。诺贝尔文学奖得主彼得·密达沃（Peter Medawar）就曾把他的一本书定名为《可解的艺术》（*The Art of the Soluble*，1967），并在书中指出，科学的一部分创造力就在于寻找处于人类知识最前沿、并可以用实证技术加以解决的问题。

心理学本身就提供了许多从无解到可解的好例子。有许多问题，诸如“一个孩子如何获得其父母的语言？”，“为什么我们会忘记我们曾经知道的事情？”，“身处一个群体中会如何改变一个人的行为和思想呢？”等，在人们意识到可以以实证的方法来解答之前的几个世纪里，都只能被猜想而已。随着这一认识的慢慢发展，心理学逐渐集合了来自各个领域中关于行为的各种问题。心理学科逐渐脱离哲学，并成为了一门独立

的实证科学。

认知心理学家史蒂芬·平克尔（Stephen Pinker，1997）讨论了“未知”可以划分为“问题”或“玄谜”。如果是“问题”，我们知道其答案是能找到的，即使我们目前还没有答案，我们也知道它大概是什么样子。如果是“玄谜”，我们甚至不能想象答案可能会是什么样子。利用这些术语，我们可以看到，科学就是将玄谜变为问题的过程。事实上，平克尔（1997）指出，他之所以要写《头脑如何工作》（*How the Mind Works*）这本书，正是因为“从心理表象到浪漫的爱情，几十个心理和思维方面的玄谜最近已经升级为问题了”（p. 9）。

心理学和世俗智慧："常识"的误区

我们每个人都有一套固有的行为模型，这些模型影响我们的交往以及我们如何看待自己和他人。事实上，一些社会、人格和认知心理学家正在探究这些内隐的心理学理论的本质。我们很少会清晰并有逻辑地表达我们的理论。相反，我们通常只有在特意关注它们或者发现它们受到某种挑战时，才会意识到它们的存在。其实，我们个人的行为模型并不像真正的理论那样具有内部一致性，相反，当我们觉得需要对行为作出解释时，往往搬出一箩筐关于人类行为的普遍真理、说教及谚语。这些关于行为的常识存在一个问题，它们之中有不少是自相矛盾的，因此也是不可证伪的（证伪原则是下一章的主题）。

人们爱用一些民间谚语来解释行为事件，即使之前在解释同一类型的事件时曾用过与之完全矛盾的谚语。例如，我们中的大多数人都听到或说过“三思而后行”——若不是我依稀记得之前有人告诫说“该出手时就出手”，我还会觉得这是个有用的、直接的行为建议呢。“小别胜新婚”明确预测了一种对于事件的情绪反应，但“眼不见，心不烦”不也同样如此吗？如果“欲速则不达”，为什么又说“时不我待”？既然“三个臭皮匠，顶个诸葛亮”，为什么又说“三个和尚没水吃”？如果我认为“行

走江湖，安全第一”，为什么也相信“不入虎穴，焉得虎子”？如果“异性相吸”，为什么又“物以类聚”？我劝许多学生“今日事今日毕”，但我希望没跟我刚刚指导过的那个学生说过这番话，因为我刚还跟他说“要顺其自然”。

这类谚语和俗话构成了对行为的固有“解释”，人们爱用它们，就是因为它们难以驳倒。不管发生什么事，都可以拿一条出来解释一番。难怪我们都认为自己是判断他人行为和人格的高手。天底下发生的事我们都能解释。

所以，有时我们固有的心理理论不容反驳。我们将在下一章中看到为什么这种不可反驳性造成了理论的失效。然而，即使我们的世俗观念有一些特定用处，甚至是经验可证的，也会产生问题。问题在于，心理学研究表明，在接受实证检验后，许多关于行为的普遍文化信念都被证明是错误的。举个例子，“工作经验对于高中生是有价值的”是一个普遍信念。大多数美国成年人认为，青少年边打工边上学是一件好事，因为：（1）他们能赚钱，支付自己未来的学费及家庭开销；（2）他们能建立“职业道德”，使得他们在将来能够成为更具责任感的员工；（3）他们会对我们的经济体系产生更大的尊重；（4）由于已经融入了经济体系，他们会成为更有动力的学生。

发展心理学家们就打工对中学生行为、态度及学业成绩的影响进行了广泛研究（Sternberg，Brown，& Dornbusch，1996）。他们发现，基本上我们所有关于青少年打工的文化信念都是错误的。一个十几岁的孩子打工所赚的钱中，只有极少量用于帮助家庭开支或继续学习。绝大多数收入花在了炫富或满足电视广告制造出的“需求”上。上学打工会对学生的教育及教育体验造成损害。最有趣的是，工作经验使青少年更加玩世不恭，不尊重工作及其在经济体系中的价值。在一项研究中，打工的青少年比不打工的青少年更可能赞同“疯子才会玩命工作”和“天下没有哪家公司会关心员工”等言论（Greenberg & Sternberg，1986）。最后，在回顾了这方面的研究之后，研究者得出结论：“打工似乎有可能引发而不是阻止了某些形式的不正当行为”（p. 6）。看来，我们炮制

了大量有关青少年打工价值的文化神话。“塑造品格”、“培养金钱观念”这类说辞都是虚假的。这类陈词滥调其实属于民间传说，是人类学家在欠发达国家从事研究时所感兴趣的——这些传说使我们感觉良好，也使当前文化习俗变得合理，但没有任何事实依据。

世俗观念（或称“常识”）出现谬误的例子俯拾皆是。比如说，有一种说法是，学习好或读书多的孩子都不擅长交际和体育。这个观点虽然错得离谱，但在当今社会上极为流行。有大量证据表明，与“常识”世俗观念相反，爱读书的人和追求学术成就者与不读书者相比，有着更强健的体魄，而且更常参与社交活动（Gage & Berliner，1984，pp. 18-19；Zill & Winglee，1990）。再比如，学习成绩好的儿童比学习成绩差的更容易被同伴接纳。读书多的人比不读书者更愿意运动、慢跑、露营、远足、维修汽车等。

许多关于行为的世俗观念一经产生便生生不息。例如，20 世纪 90 年代风行于社会和学校的一个世俗观念是，低自尊导致攻击行为。但实证研究显示，攻击行为和低自尊并无关联。相反，攻击行为似乎往往与高自尊相关（Baumeister，Bushman，& Campbell，2000）。同样，上世纪 90 年代有一个非常流行的假说认为，低自尊导致学习成绩出现问题。事实上，自尊和学习成绩之间的真正关系可能与教育工作者和家长的假设恰恰相反，是在校成绩（以及生活的其他方面）的优秀导致了高自尊，而非后者引起了前者。（Baumeister，Campbell，Krueger，& Vohs，2003；Stout，2000）。

雷德福（Redford，1999）讨论了“人类只使用了其大脑的百分之十”这一民间信念。尽管完全缺乏认知神经科学的支持（见 Beyerstein，1999；Higbee & Clay，1998），这一观点已存在了几十年，且俨然已成为所谓的“心理学公理”。雷德福援引了专栏作家罗伯特·萨缪尔森（Robert Samuelson）对“心理学公理”的定义：“虽然没有确凿的证据支持，却因不断的重复、改变了我们体验生活的方式，而被当成事实的那些信念”（p. 53）。

世俗观念并不总是不受事实的制约。有时，当与之矛盾的事实广为

人知时，世俗心理学（“常识”）也会改变。例如，几年前，一个广为流传的有关儿童的俗语是“熟得早，衰得快”（Fancher，1985，p. 141）。这条俗语反映了“童年早熟与成年异常存在关联”这一信念，这一信念得到了许多“小时神童，长大害人”的例子的支持。但在这件事上，心理学证据证明上述俗语不甚准确，这一结论已被大众文化所吸收，所以你以后几乎不大会再听到更多这样的世俗“智慧”了。

最后这个例子是一个警告，提醒我们注意今日的“常识”，因为不难看出，昨日的常识往往变成今天的谬论。毕竟，常识就是“尽人皆知的知识”，对吧？对。那么，妇女不能投票，对吧？非裔美国人不应该接受教育，对吧？残障人士不该在社会里出现而应当被送到收容所去，对吧？事实上，150 年前，这些观念都是尽人皆知的常识。当然，我们现在视这些过去的常识为谬论，都是些以完全未经证实的假设为基础的信念。但是，从这些例子中，我们可以看到心理学在常识面前扮演的关键角色。常识总是基于一些假设，而心理学对这些假设的经验基础进行检验。正如我们之前看到的许多例子，有时候假设得不到实证支持。这样的例子还有很多，通过它们，我们可以看到，心理学扮演着一种世俗智慧检验者的角色，常常难免和诸多根深蒂固的文化信念发生冲突。心理学往往是“传递坏消息的信使”，宣告原本为人们所接受的世俗观念再无立足之地。这就不难理解，为什么许多人不仅无视这些消息，还想消灭这些信使。

心理学是一门年轻的科学

建立在实证基础上的心理学始终存在反对意见。仅仅 100 年前，剑桥大学还拒绝建立一个心理物理学实验室，因为这样的一个主题研究，“以把人类的灵魂放在天平上的方式侮辱了宗教信仰”（Hearst，1979，p. 7）。心理学致力于证明其问题是实证可解的，这一战斗也是最近才取得胜利。不过随着科学的进步，心理学家将涉足越来越多的

主题，这些主题涉及人类某些牢固的信念，而很多都是可以通过实证方法验证的。心理学家现在研究的主题包括道德推理的发展、浪漫之爱、种族偏见的性质以及宗教信仰的心理和社会决定因素等。童年期性行为的研究最近引发了很多争议（Hagan，2001；Rind，Tromovitch，& Bauserman，2001）。有些人反对对这些领域进行实证调查（Hunt，1999），但这些领域都取得了科学进展。

莱文和奥唐奈（Levin & O'Donnell，2000）指出，有些人反对某些心理学研究，只因为他们认为这些领域是“不必去知道的”。他们举了一个例子：某学校的董事会给予家长两种选择，一种是让他们的孩子在K-2不同年龄混合班学习，另一种是在按年龄分的常规班中学习。教师提出对这两种方法进行研究的建议，但此建议被校董事会否决，因为他们觉得，如果一旦研究发现某种方法更有效，家长会迫使他们完全改用那种类型的教学。就像莱文和奥唐奈（2000）说的那样，“学校董事会压根儿不想知道！”（p. 66）。因此，我们应当意识到，心理学研究往往受到诋毁，不是因为人们认为它不好，而是因为他们希望避免研究结果可能给他们带来的影响。

心理学总是处于这种两难境地。一方面，一些人反对把心理学称为科学，否认心理学家可以建立关于行为的实证理论。另一方面，另一些人则由于惧怕心理学在某些行为领域揭示的真相会威胁到他们的信仰，而反对心理学家在这些领域进行研究。斯金纳学派的心理学家就总是面对这类相互矛盾的指责。例如，有批评者认为行为主义的强化律不适用于人类行为。同时，另一些批评者则担心人们会运用这些规律去对人类进行严酷的、不人道的控制。因此，行为主义者腹背受敌，一些批评者否认行为主义者所发现的行为定律有用，而另一些批评者则害怕这些定律被滥用！

上述现象的产生主要是由于，年轻的心理科学刚刚开始揭示行为方面的一些事实，而在过去，这些问题是游离于研究之外的。它的青涩多多少少也解释了为什么许多人总是对这一学科产生误解。但无论如何，在过去的40年里，心理学已经在我们称之为科学的这个相互关联的知

识体系中站稳了脚跟。认识不到这一点，就会对心理学产生各种各样的误解。

小　结

心理学是一个主题非常广泛，但又相对松散的学科，它包含一些通常不被归入同一概念的众多研究主题，然而，它们都使用科学方法来理解行为，从而实现了学科的统一。科学方法绝非是指一套生硬的规则，而是指一些非常普遍的原则。最重要的三点是：（1）科学采用系统的实证主义的研究方法；（2）它以可公开验证的知识为研究对象；（3）它研究实证可解的问题，并产生可检验的理论（下一章的主要内容）。构成系统实证主义基础的结构化及可控制的观察是本书随后几个章节的主题。科学通过同行评审等程序和重复验证等机制来保证知识的公共性。

心理学是一门新兴的科学，因而经常会和世俗智慧相冲突。这种冲突是任何新兴学科都会遇到的，了解这种冲突有助于我们理解为什么有人反对将心理学视为一门科学并对心理学持敌意态度。同时，与世俗常识之间的碰撞也令心理学成为一门激动人心的学科。很多人进入这一领域正是因为它提供了一个机会，让人们能够检验那些被毫无争议地接受了数百年的“常识”。

2

可证伪性

——如何捕捉头脑中的小精灵

1793 年，一场严重的流行病——黄热病袭击了费城。当时，这座城市里有一位顶尖的医生名叫本杰明·拉什（Benjamin Rush），他是独立宣言的签署人之一。在灾难过程中，拉什是少数几位确实治疗了几千例黄热病的医生。拉什信奉一种医学理论，认为黄热病必须用大量放血的方法治疗（用手术刀或水蛭吸血的方法使血液离开身体）。他为许多病人实施了这种疗法，当他自己感染这种疾病的时候，他也如法炮制。评论家指责他的治疗方法甚至比疾病本身更危险。然而，随着疾病的流行，拉什对他的疗法却更加自信了，即便曾有几个病人死去。这是为什么呢？

有人这么总结拉什的态度：“一方面坚信自己的理论是正确的，另一方面又缺乏有效的方法对治疗效果进行系统研究，因此他将每个好转的病例都归为治疗方法的功效，而将每个死亡的病例都归为病情的严重性”（Eisenberg，1977，p. 1106）。换句话说，如果病人情况好转，就被作为放血疗法有效的证据；如果病人死掉了，就被拉什解释为病人已经病入膏肓，无药可救。我们现在知道为什么对拉什的批评是正确的了：他的治疗方法和黄热病本身一样危险。在本章中，我们将要讨论拉什错

在哪里。他的错误为阐明科学思维中最重要的一项原则提供了样本，而这一原则在评估心理学理论时尤其有用。

本章中，我们关注第 1 章中已经讨论过的科学的第三个基本特征：科学只研究可解的问题。科学家们所说的“可解的问题”通常是指“可检验的理论”。科学家要确认某个理论是不是可检验的，采取的方法就是确保该理论是可证伪的，也就是说，理论对应着自然世界中的真实事件。接下来，我们就要看一看为何所谓的可证伪性标准在心理学中如此重要。

理论和可证伪性标准

本杰明·拉什在评估其疗法的效果时跌入了一个致命的陷阱。他的评价方法根本就不可能让人得出其治疗方法无效的结论。如果说，病人的恢复是对他治疗方法有效性的肯定（对其医疗理论的肯定），那只有当病人的死亡是对其治疗方法的否定时才算公平。但事实上，他却把这种否定合理化了。拉什解释证据的方式，违反了科学理论建构和检验应遵循的最重要原则之一：他令自己的理论不能被证伪。

科学理论的表述应该遵循这样的原则——从中得出的预测有可能被表明是错误的。因此，对某理论的新证据进行评价，必须使新的数据具有证伪该理论的可能性。这项原则通常被称为“可证伪性标准”。一位叫卡尔·波普尔（Karl Popper）的哲学家一直致力于强调可证伪性标准在科学进程中的重要作用，他的文章被现在仍从事科研工作的科学家们广泛阅读（Magee，1985）。

可证伪性标准主张，一项理论如果有用，它所做出的预测必须是明确的。理论必须两面兼顾，也可以说，这项理论在告诉我们哪些事情会发生的同时，应该指出哪些事情不会发生。如果不会发生的事情确实发生了，我们就得到了一个明确的信号——这项理论有问题：它可能需要修正，或者我们需要去寻找一个全新的理论。不管哪种方式，我们将最

终有一个更接近真理的理论。相反，如果一项理论预测包括了所有可能观察到的数据，那么它将永远不能被修正，同时我们将被禁锢在当前的思维方式中，失去了取得进步的可能。这就是说，一项成功的理论并不是可以用来解释所有可能的结果，因为这样的理论本身就丧失了任何预测能力。

在这本书的余下部分，我们会经常涉及理论的评估，因此我们必须澄清一个关于理论的常见误解。这个误解体现为我们常说的一句话:“哦，这只不过是一种理论。”这句话代表了外行人使用“理论”这个词时通常所指的意思：一项未经证实的假设，一个纯粹的猜想或直觉。这意味着一个理论与其他理论并无优劣之分。“理论”这个词在科学上绝对不是这么用的。当科学家说到“理论”的时候，他们指的不是未经验证的猜想。

科学上的理论是一组具有内在联系的概念，它们能对一组数据做出解释，并对未来实验的结果做出预测。假设是从理论中产生的具体预测（理论则更加普遍和全面）。目前可行的理论是那些产生了一些假设，并且其中许多已经得到了验证的理论。因此这种理论的理论结构与大量的实证观察相一致。然而，当观察数据开始与理论中提出的假说相矛盾的时候，科学家们会尝试构建一个能为数据提供更好解释的新理论（或者，在更通常的情况下，只是修正已有的理论）。因此，目前在科学范畴内所讨论的，都是在一定程度上已经被证实了的、所做出的预测并没有与现有的数据相矛盾的理论。它们并非纯粹的猜想和直觉。

外行人和科学家们使用“理论”这个词时的这种差异，经常会被一些试图将神创论纳入公立学校教育的虔诚的正统基督教徒所利用（Forrest & Gross，2004；Scott，2005；Talbot，2005）。他们的论点通常是“进化论毕竟只是理论”。这种观点试图借用外行人对“理论”术语的用法，蓄意将理论歪曲为“只是一个猜想”。然而，通过自然选择的进化理论不是外行人所理解的“理论”（相反，在外行人的理解中，它应被称之为“事实”，见 Randall，2005），而是一个科学意义上的理论，是由一系列庞大而多样的数据支持的结论（Maynard Smith，1998；

Ridley，1996，1999；Scott，2005）。它并不等同于其他任何猜想，不是一个纯粹的猜测。相反，它与从属于其他学科的知识紧密相联，这些学科包括地质学、物理学、化学以及生物学的各个分支。著名的生物学家西欧都萨斯·杜赞斯基（Theodosius Dobzhansky）（1973）在他的一篇题为《生物学中除了进化论以外，别无他物》（Nothing in Biology Makes Sense Except in the Light of Evolution）的著名文章里就阐述了这一观点。

敲门节奏理论

下面假设一个例子来展示可证伪性标准是如何起作用的。一个学生在敲我的门。跟我同一办公室的同事有一套“不同的人以不同的节奏敲门”的理论。在我开门之前，我的同事预言门后是一位女性。我打开门，这个学生确实是女的。事后我告诉同事，他的表现令我惊叹，但这种惊叹程度非常有限，因为，即使没有他所谓的“敲门节奏理论”，他也有50% 的正确几率。他说他的预测能高于随机水平。另一个人来敲门，我的同事预测说，这是个男性，而且不到 22 岁。我打开门，果然是个男生，而且我知道他刚从中学毕业。我承认我有点被震撼了，因为我所在的大学有相当数量的学生是大于 22 岁的。当然，我仍然坚持说校园里年轻的男性相当普遍。见我如此难以被取悦，我的同事提出做最后一次测试。在下一个人敲门之后，我的同事预测：女性，30 岁，5 英尺 2 英寸高，左手拿书和挎包，用右手敲的门。打开门后，事实完全证明了预测，对此我的反应截然不同了。我不得不说，如果我的同事不是使用诡计事先安排这些人出现在我门口的话，我现在的确非常震惊。

为什么我的反应会不同呢？为什么我同事的三次预言会让我产生三种不同的从“那又怎么样？”到“哇哦！”的反应？答案与预测的具体性和精细度有关。越精细的预测在被证实的时候会给我们越大的触动。要注意，不管怎样，精细度的变化和可证伪性直接关联。预测越具体和精细，有可能证伪它的观测现象就越多。例如，有很多不是 30 岁和 5

英尺2英寸高的女性。请注意这里的暗示：从我截然不同的反应可以看出，一个能够预测出最多不可能事件的理论最容易将我征服。

好的理论做出的预测总是会显示自己是可证伪的。坏的理论不会以这种方式把自己置于危险的境地，它们做出的预测是如此笼统，以至于总会被证明为正确的（例如，下一个来敲我门的人会是100岁以下），或者，这些预测会采用一种能免于被证伪的措辞方式（如本杰明·拉什的例子）。事实上，当一种理论被置于“不可被证伪”的保护下，那么可以说它已经不再是科学了。事实上，哲学家卡尔·波普尔正是由于试图界定科学和非科学的区分标准，才会如此强调证伪原则的重要性。这里的讨论和第1章中我们有关弗洛伊德的讨论，甚至与心理学之间都有直接的联系。

弗洛伊德与可证伪性

在本世纪最初的几十年，波普尔一直在探寻，为何一些科学理论似乎导致知识的进步，而其他一些则导致智力停滞（Hacohen，2000）。例如，爱因斯坦的广义相对论引发了一系列惊人的发现（例如，从一个遥远的恒星发出的光线经过太阳附近时发生弯曲），恰恰是因为它是这样建构预测的：许多事件或现象一旦被证实与之相矛盾，就可以证伪该理论。

波普尔指出，一些使知识停滞的理论却并非如此，并以弗洛伊德的精神分析法作为例子。弗洛伊德的理论使用一个复杂的概念结构，在事后解释人类行为，但并不做事前的预测。它可以解释一切，但是波普尔认为，也正是这个属性使得它在科学上无用。它不做具体的预测。精神分析理论的拥护者花费大量的时间和精力试图用他们的理论解释人类所有已知的活动——从个人的怪癖行为到广泛的社会现象，但他们在使这个理论成功地成为事后解释的丰厚资源时，也剥夺了其所有的科学实用性。如今，弗洛伊德的精神分析理论在激发文学想象方面比在当代心理学中扮演着更重要的角色（Robins，Gosling，& Craik，1999，2000）。

它在心理学中的地位日益下滑，部分原因就是未能满足可证伪性标准。

这种不可证伪理论的存在会导致实际的危害。正如一位评论家所指出的："不正确的但被广泛传播的有关心理的理念，不可避免地会对社会造成危害。由于精神分析学派的声望一度被人为地抬高，令社会上许多有疾病及遗传缺陷的人拒绝有效的治疗，转而从个体早期经历中寻找自己现有痛苦的根源"（Crews，1993，p. 65）。以抽动性秽语症为例。这是一种以身体抽搐和痉挛为特征的紊乱，并伴有言语症状，如嘟囔、吠叫、模仿言语（无意识地重复他人的话）和秽语（强迫性重复淫秽词语）。抽动性秽语症是一种器质性的中枢神经系统紊乱，并已经成功地被药物治疗所攻克（Bower，1990，1996a）。纵观历史，抽动性秽语症患者一直遭受着迫害，早期被宗教统治者视为妖魔，近代又被认为是鬼怪附体，要被强制驱魔（Hines，2003）。更重要的是，在 1921 至 1955 年之间，对这种病的解释及疗法一直被精神分析学派的概念体系所把持，这在很大程度上阻碍了人们对此病成因及治疗的理解（见 Kushner，1999）。有关这种病症的不可证伪的精神分析解释层出不穷。这些似是而非的解释所造就的概念泥潭蒙蔽了这一病症的实质，也阻碍了对其进一步的科学探究。例如，有一位作者曾经这样写道：

> （抽动性秽语症是）精神分析导致脑部疾病研究发生倒退的典型例子。勒·图雷特（La Tourette）将疾病归因于大脑的退行性变化过程。而在本世纪最初的几十年，由于弗洛伊德理论的盛行，对这种病的关注偏离了大脑……这一倒退的结果使病人往往被转到精神科医生（通常是精神分析学派的医生）而非神经科医生那里，因此没有接受生理检查和研究（Thornton，1986，p. 210）。

夏皮罗等人（Shapiro et al.，1978）提到，一位精神分析师认为，他的病人"不愿意放弃抽动，因为这成了她性快感的源泉和潜意识性欲的表达"。另一位精神分析师则认为，抽搐"等同于手淫……与生殖器快感相联系的力比多转移到了身体的其他部位"。第三位认为抽搐是一种"肛门施虐的迁移症状"。第四位认为，抽动性秽语症的患者具有"强

迫型人格以及自恋倾向”，病人的抽动“代表了一种情感症状，对想表达情感的压抑性防御”。夏皮罗等人（1978）对这类理论现状的总结，很好地说明了忽视可证伪性标准的有害影响：

> 精神分析这种理论化的方式简直面面俱到。抽搐是迁移性的症状而非歇斯底里症、肛门的而又是性欲的、受意志控制的而又是强迫性的、器质性病变同时又与原始心理动力有关……这些心理标签、诊断和治疗被不幸地强加在病人及其家属身上，而且是以一种毫不谦卑、相当武断、伤害巨大的方式。因为其随后的广泛影响，这些观点为对此病症的认识和诊治造成了极大的障碍。（pp. 39–42，50，63）

当研究人员承认精神分析的“解释”对治疗该疾病毫无用处的时候，对抽动性秽语症的认识和治疗才开始获得进展。那些毫无用处的解释是诱人的，因为它们似乎能对事情进行解释。事实上，它们都是在事后对所有事情做出解释的。然而，它们提供的解释不过是制造了理解的幻觉。由于总试图在事后解释一切，它们也就堵死了前进的大门。只有当一种理论并不预测所有事情，而是提出具体的预测——提前告诉我们哪个特定的情形会出现时，该理论才会出现进步。当然，从这样的理论推导出的预测可能是错误的，但这是优势，而非缺点。

小精灵

如果人们能够从所研究的问题里跳出来，尤其是，如果人们能以史为鉴的话（如本杰明·拉什的例子），就不难识别出那些不能证伪的概念体系。当其例证明显是编造的时候，也很容易察觉其不可证伪性。举例来说，大家还不知道，我已经发现有一种大脑机制在控制行为，你很快就会在随处可见的八卦杂志上看到这个发现。我发现在大脑左半球的语言区附近住着两个小精灵，它们有能力控制大脑许多区域中的电化学过程。而且，长话短说，它们基本上控制了一切事情。但是，有一个问

题阻止我们看到它们，那就是小精灵有能力发现任何对大脑的侵入（外科手术、X 光等），一旦觉察到外界的探测，它们就会消失（我忘记说了，它们具备隐身能力）。

毫无疑问，我在这里是用一个更适合小学生的例子来侮辱你的智慧。很明显这个例子是我捏造的，但我对小精灵的假设永远无法被证实是错误的。然而，考虑一下。作为心理学导论的讲师和公开演讲者，我经常被问到，为什么不讲授在过去几年里在超感官知觉（ESP）和通灵学方面取得的那些惊人的新发现。我不得不告诉这些提问者，他们所获悉的大多数相关信息，无疑都是来自于大众媒体，而非科学界所承认的信息来源。事实上，一些科学家曾关注过这类说法，但没能够重复这些发现。我要提醒各位读者，要将一个研究成果认定为确定的科学事实，可重复性是至关重要的，尤其是当研究结果与以前的数据或现有的理论相矛盾的时候。

我甚至可以坦率地说，许多科学家对 ESP 研究已经失去了耐心。原因当然与此领域充斥着欺诈、江湖骗术和媒体炒作有关，但令科学界觉醒的更重要的原因是马丁·加德纳（Martin Gardner，1972）所谓的“ESP 研究第 22 条军规”。

其运作方式如下：一名“信奉者”（在开始调查之前就相信 ESP 现象存在的人）声称已在实验室证明了 ESP。一名“怀疑者”（质疑 ESP 存在的人）被邀请证实这种现象。通常，在观察实验情境之后，怀疑者会要求信奉者进行更多的控制（我们会在第 6 章中讨论这种类型的控制），虽然这些要求有时候会被拒绝，但通常善意的信奉者们会同意他们的要求。当加入了实验控制之后，这种现象就不再出现了（见 Alcock，1990；Hines，2003；Humphrey，1996；Hyman，1992，1996；Kelly，2005；Marks，2001；Milton & Wiseman，1999）。怀疑者会对这种失败做出正确的解释——早先对这个现象的证实是由于缺乏足够的实验控制，因此结论不能被接受。但他们往往吃惊地发现，信奉者并不承认早先的证明是无效的。相反，他们搬出超感知的“第 22 条军规”：他们坚称，心理能量是很敏感的、微妙的，并容易受到影响。怀疑者

的“负面感应”是瓦解这一“超感官能量”的罪魁祸首。信奉者认为，怀疑者的“负面气场”被移开后，这种心理能量无疑会回归。

这种对无法在实验室中证实ESP的解释方式，在逻辑上与我编造的小精灵的故事相似。ESP的运作就像小精灵一样。只要你不侵入性地仔细观察它，它就在那儿。如果你观察它了，它就不见了。如果我们接受这种解释，那么向怀疑者证明这一现象就变得不再可能。这种现象只为信奉者现身。当然，这种说法在科学领域是不能接受的。我们没有磁力物理学家和非磁力物理学家之分（即磁场只对前者存在）。以这种方式解释ESP的实验，使得ESP的假设变得像小精灵的假设一样不可证伪。正是这种解释方式，将ESP排除在了科学殿堂之外。

不是所有的证实都等价

可证伪性原则对于我们如何看待一个理论的证实过程具有重要的意义。许多人认为，一个好的科学理论就是被多次证实的理论。他们假设，被证实的次数是对理论进行评价的关键。但是，可证伪性原则意指，理论被证实的次数并不是最重要的因素。原因在于，正如“敲门节奏理论”所展示的那样，并不是所有的证实都是等价的。证实能否令人信服，取决于预测在何种程度上将自己暴露在可能被证伪的情境下。一个非常具体的、可能被证伪的预测（例如，一位女士，30岁，5英尺2英寸高，左手拿书和挎包，用右手敲门），比20个不可证伪的预测（例如，一个小于100岁的人）拥有更强的说服力。

因此，我们不能仅关注理论被证实的数量，更要关注验证本身的质量。将可证伪性作为一种评价标准，就可以使那些使用研究结果的人抵制不科学的、全能理论的诱惑。这种全能理论会不可避免地妨碍我们对世界和人类本质进行更深入的探索。事实上，这种理论的死角也正是最魅惑人的地方，因为它们永远不能被证伪。在纷繁多变的现代世界中，这种理论千年不变。

波普尔经常指出，“这些（不可证伪的）理论拥有巨大的心理吸

引力，其秘密在于它们能够解释一切事情。预先知道无论什么事情发生，你都能理解它，不仅给你智力上的掌控感，而且，更重要的是，让你拥有应对这个世界所需的安全感”（Magee，1985，p. 43）。但是，这种安全感的获得并不是科学的目标，因为对这种安全感的追求是以知识发展的停滞为代价的。科学是一套不断挑战原有信念的机制，在这种机制里，原有信念以一种能够被证伪的方式接受实证检验。这一特点往往使科学（尤其是心理学）与所谓的世俗智慧或者常识直接发生冲突（正如我们在第 1 章中所讨论的）。

可证伪性和世俗智慧

心理学威胁到世俗智慧所提供的安逸感，因为作为一门科学，它不能只提供无法被反驳的解释。心理学的目标是对各种行为理论逐一进行实证检验和筛选。某些世俗智慧表述得很清晰，经得起实证检验，这当然是心理学所欢迎的，而且其中许多已经被纳入了心理学理论。然而，心理学并不追求那类事后能解释一切，但事先无法做出任何预测的理论，不追求这种解释系统所带来的安逸感。它不接受那些被设计得永不可变、并代代相传的世俗智慧体系。试图向学生和公众隐瞒这一点无疑是自毁长城。不幸的是，一些心理学指导教师和普及者觉察到了心理学对世俗智慧的威胁给一些人造成的困扰，于是他们有时会通过传递错误信息来试图安抚这种情绪，如“你会学到一些有趣的东西，但别担心，心理学不会挑战那些你深信不疑的观点”。这是一个错误，它对“什么是科学”和“什么是心理学”都会造成混乱。

科学寻求概念上的变化。科学家试图描绘世界的真实图景，这个图景可能与我们的固有信念正好相反。现代思潮中有一种危险的趋向——认为应避免让一般大众知道世界的真正本质，一种无知的面纱是必要的，以防公众面对真相时手足无措。心理学与其他科学一样，拒绝向人类隐瞒真相的观点。生物学家米希尔·吉瑟林（Michael Ghiselin）进一步宣称，当知识不能广泛普及时，我们都会蒙受损失：

> 如果我们有健康的邻居，那么我们会过得更好，而通过垄断药品的供应使自己变得比他们更健康的做法是愚蠢的。知识也是如此。我们邻居的无知和不健康对我们一样有害，而无知很可能正是造成他们健康问题的原因。工业的进步和我们所有人都受益于技术性劳动力的供给，我们依赖于他人的技能和专门知识。

同吉瑟林一样，心理学家也认为，当我们被那些对人类行为有误解的人们所包围的时候，大家都会蒙受损失。公众对于教育、犯罪、健康、生产力、儿童福利和许多其他重要问题的态度塑造了我们的世界。如果这些态度源于错误的行为理论，那么我们大家都会受到伤害。

承认错误的自由

科学家们发现，可证伪性原则的一个最具解放意义和最有用的启示是：在科学上，犯错并不是罪过。被证伪的假设为科学家们提供了信息，他们可以用来调整理论，使理论更紧密地与数据相一致。哲学家丹尼尔·丹尼特（Daniel Dennett，1995）曾说过，科学的本质就是“在公众面前犯错”（p. 380）。当数据与理论不符时，通过对理论进行不断地修正，科学家们最终构建起能更好地反映世界本质的理论。

事实上，如果我们能够在日常生活中使用可证伪性原则的话，我们生活的质量将会大大改善。这就是为什么我在本节的第一句话中使用“具有解放意义”这个词的原因。它包含着一种个人化的期许，即此处产生的理念能够同时对科学之外的领域有所启示。如果我们能够理解这一切，当我们的信仰与观察到的事实相冲突时，我们最好是调整信仰而不是否认事实和坚持错误的想法，这样我们将会少遇到一些个人和社会问题。物理学家罗伯特·奥本海默认为：

> 每个人在训练自己的推理能力时都要学会这一点，当他回顾过去时会说“这个问题我当时没想明白”。生活中的人们也需要这种能力，即脸不红心不跳地说：“我当时搞错了。”而科学能让认错每

> 时每刻都发生。你注意到某些东西与你已经思考了很久的事情相冲突或不一致，它会触动你改变想法。这与人们没完没了地去寻找一些合理说法来为自己先前的错误辩解的做法完全不同（Dos Passos，1964，pp. 150–151）。

当你与某人激烈地争论的时候——也许就是当你给出一个有力的反击来捍卫你的观点的时候——有多少次你会突然意识到你搞错了某个关键事实或论据？这时你会怎么做？你会收回前面的话并向别人承认错误，同时承认别人的解释现在看起来比你的更合理吗？或许不会。如果你和我们中的大多数人一样，那么你一定会"没完没了地寻找一些合理说法来为自己先前的错误辩解"。你试图在拒不承认失败的情况下使自己从争论中全身而退。你最不可能做的就是承认自己错了。这样的话，你和争论对手都会更加疑惑：到底哪一种信念更接近真理？如果争论不能成为公共性的（如在科学中那样），如果正确和错误的信念以同样激烈的方式争辩，如果争论的结果不能得到正确的反馈（如本例），那就没有更可靠的机制使信念与现实相吻合了。这就是为什么那么多私人和公开的对话令人困惑，为什么相比所谓的常识或世俗智慧，心理科学在解释人类行为的原因方面更加可靠。

在科学中犯错是正常的，对于科学进步来说，真正的危险是人类有一种固有的倾向，即避免让其信仰暴露在可能被证明是错误的情境之中。许多科学家已经证实了这一观点的重要性。诺贝尔奖得主彼得·米德瓦（Peter Medawar，1979）写道：

> 在能够被更好的假设所取代的情况下，错误的假设尽管是情有可原的，但它们确实会对曾经深信它们的科学家造成严重的伤害。这些人深深地热爱他们的假设，因而不愿意接受否定的实验结果。有时，他们不让自己的理论假设接受严格的检验，而是绕过这类检验，仅仅让假设中无足轻重的部分接受检验，或不断寻求一些边缘化的证据来支持假设，而不愿冒可能被彻底推翻的风险。对于任何时代的任何科学家，没有比这再好的忠告了：**一个假设在何种程度**

上被确信为正确实际上与其是否为真无关。（p. 39；原文为意大利语）

心理学界许多最具声望的科学家都遵循米德瓦的建议。在一篇报道实验心理学家罗伯特·克诺德（Robert Crowder）职业生涯的文章中，引述了他的一位同事马扎林·巴纳吉（Mahzarin Banaji）的话：“他是我认识的最不维护自己理论的科学家。如果你发现一种方法证明他的理论有漏洞，或者他的实验发现有局限性或有缺陷，他会非常高兴，并和你一起计划如何推翻该理论”（Azar，1999，p. 18）。艾泽（Azar，1999）描述了克诺德如何提出了一个叫做“前分类听觉存储器”的记忆成分理论，然后又仔细地设计了一个实验研究证伪了自己的模型。最后，进化心理学家约翰·图比（John Tooby，2002）在一篇精彩的评论中提到是何种态度使达尔文对科学做出不朽贡献：“达尔文比他同时代的人走得更远，因为他并没有受缚于‘让宇宙符合自己预期’的冲动”（p. 12）。哲学家乔纳森·阿德勒（Jonathan Adler，1998）则采用了另一种说法：“一个真正开明的人愿意跟随证据的引导。开明的人愿意听从公正的调查，而非他自己的预测。科学方法是对世界的印证，而不是对我们自己的印证”（p. 44）。

但是，要让科学发挥作用，并不需要每位从事科学工作的科学家都具备证伪的态度。雅各布·布罗诺夫斯基（Jacob Bronowski，1973，1977）在他的许多文章中指出，科学那种揭示世界真知的独特力量，并不产生于科学家们独特的德行（即他们是完全客观的、他们在解释研究结果时从来不带偏见等）。实际上，这种力量的产生是因为会犯错的科学家们身处一个证实与平衡的程序中。在这个程序中，总会有其他科学家提出批评并发现他们同行的错误。哲学家丹尼尔·丹尼特（Daniel Dannett，2002）提出过相同的论点：不是每位科学家都必须表现出罗伯特·克诺德的客观性。正如布罗诺夫斯基和丹尼特所强调的那样，科学家和其他任何人一样容易犯错，但认识到他们及其所属团体的犯错根源之后，他们设计出精巧的系统来约束自己，努力防止自身弱点和偏见

影响自己的研究结果（p. 42）。心理学家雷·尼克尔森（Ray Nickerson, 1998）以一种更为幽默的说法道出相同的观点：科学家们的虚荣心实际上在科学进程中起着作用，“科学家对自己的想法抱有的批判性态度并没有在很大程度上导致科学的成功……更真实的情况是，每个科学家都积极地想要证明某些科学家所持有的观点是错误的”（p. 32）。这些作者认为，科学知识的力量并不是来自于科学家的德行，而是源于他们不断交叉检验彼此的知识和结论的这一社会过程。

想法不值钱

先前关于检验世俗智慧的讨论，将我们引向了可证伪性原则的另一个有趣推论：想法不值钱。说得更准确些，我们的意思是某些类别的想法不值钱。生物学家和科学作家史蒂芬·古尔德（Stephen J. Gould, 1987）对此有所阐述：

> 15 年的月刊专栏写作生涯，让我收到各个科学领域非专业读者的海量来信……我发现一个常见的、同时是压倒性的错误观点。人们会告诉我他们提出了一项革命性的理论，它会拓展科学的边界。这些理论通常以单倍行距打印在几张纸上，内容通常是对最深层的终极问题的猜测——什么是生命的本质？宇宙的起源？时间的起点？但是，这些想法不值钱。任何智力正常的人都能在早饭前想出几个这样的念头。科学家们自己也很容易就能想出来。但我们不这样做（或者说，我们只让它们留在自己脑子里），因为我们不能找到方法来验证它们，以决定它们的对错。一个既不能被证实也不能被证伪的可爱想法，对科学来说又有什么用呢？

古尔德对最后一个问题的回答是：“没有用。”古尔德这里所说的廉价想法正是我们早先在对卡尔·波普尔观点的讨论中提到的那些：包罗万象、复杂、“模糊”、能够用来解释一切的宏大理论——这种理论的建构更多是为了提供情感支持，因为它们没打算被改变或抛弃。古尔德告

诉我们，这种理论对于科学目标是无用的，无论它们多么有抚慰功能。科学是创造性的过程，但是这种创造性需要让概念结构符合实验数据。这并不容易做到。那些如实解释真实世界的想法一点儿也不廉价。也许这就是为什么好的科学理论很难提出、而不可证伪的伪科学信仰体系泛滥的原因，因为后者容易建构得多。

科学理论与世界紧密联系。它们是可证伪的，并能做出明确具体的预测。事实上，形成真实的、科学真正可以解释的理论是一项困难的任务。但是，理解科学运作的一般逻辑并没有那么困难。事实上，现在已经出版了不少专为儿童撰写的关于科学思维逻辑的书籍（Kramer，1987；Swanson，2001，2004）。

科学中的错误：逼近真理

在解释可证伪性原则的过程中，我们已经勾勒出科学进步的简单模式。提出理论、并从中推导出假设，然后假设接受各种技术或方法的检验——我们将在本书余下的部分讨论这些技术。如果假设通过了某些实验的检验，该理论就得到了某种程度的确证；如果假设被实验证伪，这个理论就得做出某种程度的改变，或者被一个新理论所取代。

当然，虽然科学上的知识是暂时性的，由理论得出的假设可能是错误的，但这并不是说所有的一切都要被拿来检验一番。科学中有很多理论已经被确认过无数次，它们被称为“公理”，因为它们几乎不可能被未来的实验推翻。我们不大可能在某一天发现，血液不是循环的，或者地球并没有在环日轨道上。这些众所周知的事实并不是我们一直在讨论的假说。它们也不是科学家们的兴趣关注点，因为它们已经是确定无疑的。科学家只对已有知识范围之外的问题感兴趣：它们并不是确定无疑的。

科学实践的这一面——科学家侧重于已知事实的前沿，而忽视那些已经被充分证实的问题（所谓的公理）——对大众来说很难理解。科学

家们似乎总是更强调未知的事物而非已知事物。这千真万确，而且科学家有很好的理由这么做。为了推进知识的进步，科学家们必须一直身处已知的前沿。当然，这里是很多事情都不确定的地方。但科学进步正是通过这个过程来实现的，即试图在已知的前沿减少不确定性。这种特点常常使得科学家被公众视为是“没谱的”。但这只是表面现象，科学家们只是对知识的前沿不确定——这使我们对于事物的理解不断加深。科学家们不怀疑那些被很多研究重复证实的事实。

同样需要强调的是，当科学家通过观察法证伪一个理论或用一个新理论代替旧理论的时候，并不意味着他们要将先前用以建立旧理论的事实全都扔到一边（我们会在第8章展开讨论这个话题）。相反，新理论应该能够解释所有旧理论能解释的事实，还能够解释旧理论不能解释的事实。理论被证伪并不意味着科学家非得建构一个全新的理论。科普作家伊萨克·阿西莫夫（Isaac Asimov）在一篇题为《错误的相对性》（*The Relativity of Wrong*，1989）的文章中很好地说明了理论修正的过程，文中谈到我们对地球形状的理解是如何完善的。他首先提醒我们，不要以为“地球是平的”这一古老信念是愚蠢的，在平原上（大部分有文字的人类文明都发源于平原），地球看上去相当平坦。阿西莫夫要求我们试着对不同的理论进行定量的比较，看结果会告诉我们什么。首先，我们能够将不同理论表述为它们预测地球表面每公里曲率的大小。“地平理论”会说每公里的曲率为0。现在我们都知道，这种理论是错误的。但从某种意义上说，它又很接近真理。正如阿西莫夫（1989）所述：

> 亚里士多德之后的一个世纪，古希腊的另一位哲学家埃拉托塞尼斯（Eratosthenes）指出，太阳在不同纬度上投射不同长度的影子（如果地球是平面的，所有的影子应该一样长）。根据影子长度的不同，他计算出地球的周长为2.5万英里，那么这个球体曲率是0.000126度/英里。正如你所见，这个数值非常接近0……这从0到0.000126的差别解释了为何我们用了如此长的时间，才放弃“地球是平的”这一观念，并转而相信地球是球状的。提醒你一下，即

使是像0~0.000126之间这样细小的差别也是至关重要的。失之毫厘，谬以千里。如果这点小差别没有被考虑到，如果地球被认为是一个平面而不是一个球，那么我们将无法精确地绘制地球上大面积区域的地图。（pp. 39−40）

当然，科学没有止步于“地球是球状的”这一理论。我们早先讨论过，科学家们一直在尝试尽量改进他们的理论，并挑战当前知识的局限。例如，牛顿的引力理论预言地球并不是完美的球形，这个预言确实被证实了。现在已经证明，地球在赤道附近略微凸起，而在两极附近略微扁平。这是个被叫做“*扁球体*”的形状。地球从北极到南极的直径是7 900英里，赤道直径是7 927英里。所以，地球的曲率并不是一个常数（像一个理想的圆球那样），而是在每英里上有约7.973英寸到8.027英寸的微小变化。正如阿西莫夫（1989）所言：“从球体到扁球体的修正比从平面到圆的修正要小得多。因此，虽然‘地球是球状的’这一理解有误，但严格地说，它没有错到‘地球是平的’那种程度。”阿西莫夫关于地球形状的例子为我们展示了科学家们使用错误、误差和证伪这些术语的不同情境。这些术语并不是说被检验的理论错得一无是处，这些理论仅仅是不完善的。所以当科学家强调说理论是暂时性的、可能被未来的研究发现所修正的时候，他们所指的就是例子当中的情形。当科学家相信地球是球状的时，他们认识到在未来某一天，这个理论需要在细节上进行修正。无论如何，从球体到扁球体的变化维持了地球是一个球体的“大体正确性”。我们绝不会在某天醒来突然发现它其实是一个立方体。

临床心理学家斯科特·利连费德（Scott Lilienfeld，2005）向心理学专业的学生介绍了阿西莫夫的观点：

当向学生解释心理学知识本来就是暂时性的、可以被修正的时候，有些学生会错误地得出结论，认为真正的知识是不存在的。这种观点在某些后现代主义圈子里非常流行，它忽视了对不同确定程度的知识的区分。虽然绝对的肯定在科学中无法实现，但一些科学理论，如达尔文的自然选择学说，已经被极好地证实了，而其他一

些理论，如支持占星术的理论，已经遭到了有力的驳斥。还有一些理论，如认知失调理论，仍处在科学争议当中。因此，科学理论是个确定程度不同的连续体：有些已经成为了确定的事实，另外一些则被完全地证伪了。对于科学问题，方法论上的怀疑主义并不产生完全确定的答案（原则上说，这些答案可能会被新的证据推翻），这个事实并不意味着知识是不存在的，只是说知识是暂时性的。

小　结

科学家们提到“可解的问题”时,通常指的是“可检验的理论”。“可检验的理论”的定义在科学上是非常明确的：这个理论是有可能被证伪的。如果一个理论不可证伪，并且和自然界的真实事件没有关联，那么它就是无用的。心理学里一直充斥着不可证伪的理论，这也正是心理学发展缓慢的原因之一。

好的理论能够做出具体的预测，具有高度的可证伪性。相比于一个不精确的预测，一个明确具体的预测如果得到证实，会为产生这个预测的理论提供更大的支持。简言之，可证伪性原则的一个含义就是，并非所有理论的验证都具有同样的价值。可证伪性越高，预测越具体，得到证实的理论就越受青睐。即使预测并没有得到证实(比如它们被证伪了)，可证伪性对于理论的发展也是有用的。一个被证伪的预测说明，原有理论要么应当抛弃，要么需要进行改变以解释不一致的数据。正是通过这种由被证伪的预测所引发的理论修正，像心理学这样的科学才能逐步向真理逼近。

3

操作主义和本质主义

——“但是，博士，这到底是什么意思？”

物理学家真正理解地心引力是什么吗？我的意思是真正。他们知道“地心引力”这个术语的真正含义是什么？它的内在本质是什么？说到地心引力时最终所要表达的意思是什么？说到底，它究竟是什么？

类似这样的问题反映了一种科学观点，哲学家波普尔称其为“本质主义”。这种观点认为：从内在本质或者本质属性的角度对现象做出最终解释，才算得上是好的科学理论。支持这种观点的人通常也相信，无法对现象做出最终解释的任何理论都是无用的，这样的理论不能反映真实的内在情况，不能反映世界存在方式的本质。本章，我们将讨论为什么科学不去回答本质主义者的问题，而是通过对概念进行操作性定义得以进步。

为什么科学家不是本质主义者

事实上，科学家并不企图获得本质主义者所追求的那类知识。从这一意义上讲，对于本章一开始提出的问题的正确回答是：科学家不知道

地心引力是什么。科学并不试图回答关于宇宙的“终极”问题，彼得·米德瓦（Peter Medawar，1984）曾写道：

> （确实存在）那些科学不能回答并且在科学发展的可预见的范围之内也不可能得到答案的问题。比如那些孩子们会提出的问题——“终极问题”……我能想到的这样的问题有：世界是如何开始的？我们来到这世间是为了什么？生活的意义是什么？（p. 66）
>
> （然而，）即使科学不能回答终极问题，但也不意味着必须接受其他的答案；也不能理所当然地认为，既然这类终极问题能被提出，就一定能够被回答。就我们目前的理解力而言，这类问题是无从回答的。（p. 60）
>
> （但是，最终）就它能回答的那类问题而言，科学的潜力是无穷的……没有什么可以阻挡或中止科学的发展，除了诸如缺乏勇气之类的道德方面的缺陷。（p. 86）

科学家之所以质疑那些自称为终极问题给出绝对答案的人、理论或者观念体系，一个原因就是科学家认为终极问题是无法回答的。科学家并不会宣称他们可以提供完美的知识；科学的独特优势并不在于它是一个不会犯错的过程，而在于它提供了一种消除错误的方式，它能不断消除我们认识中的错误。再者说，自称完美或绝对知识的主张及做法，却往往会阻碍人们的探索。自由而开放地探索知识是科学活动的一个先决条件。科学家们总是在怀疑那些号称已经找到问题最终答案的言论。

本质主义者喜欢咬文嚼字

本质主义者的态度通常有一种表现：在探求知识之前，过于关注术语或概念的定义。“但是我们必须首先界定我们的术语”是本质主义者常用的一个口号。“某理论性概念的真正含义是什么？”这种理念似乎意味着，当一个词被当做理论中的概念使用之前，我们必须对这个词的

使用所涉及的所有潜在语言问题有一个全面而且清晰的理解。事实上，这正好与科学家的工作方式相反。在对物理世界开展研究之前，物理学家不会花费气力讨论如何使用“能量”一词，或者当我们讨论物质的基本组成时，“粒子”一词是否真正表达了我们要表达的本质含义。

在科学领域里，确定某概念的意义，是在与该术语有关的现象得到一定程度的研究之后，而非研究之前。一个精确的概念性术语来自科学过程中固有的那种数据和理论间的相互作用，而不是关于语言用法的辩论。本质主义者让我们陷入无休止的文字争论，而许多科学家坚信这样的文字游戏使我们脱离了事物的实质。例如，对于“生命一词的真正含义是什么”这个问题，两个生物学家的回答是“没有什么真正的含义，它只是足够好地满足我们生物学家工作需要的一种用法，并不是争论或辩驳的主题”（Medawar & Medawar，1983，pp. 66-67）。总之，科学家的目的是解释现象，而非对措词进行分析。在所有的科学学科里，进步的关键在于放弃本质主义，接受操作主义。这正是本章中我们探讨的主题。没有别的学科比心理学更能说明这个问题了。

操作主义者将概念和可观测事件联系在一起

那么，如果不是来自于语言文字的争论，科学中概念的含义又来自于哪里呢？正确使用某一科学概念的标准是什么？为了回答这些问题，我们必须讨论操作主义。它对于在科学领域中建构理论至关重要，尤其对于评估心理学中的理论及观念具有重要作用。

尽管操作主义形式多样，但是对于科学信息的使用者来说，用最广泛的思路去思考操作主义是最有效的。“操作主义”只是这样一种思想：科学理论里的概念必须立足于可观测事件，或与可观测事件相关联，而这些可观测事件是可以被测量的。将概念与可观测事件相联系的是概念的操作性定义，这使概念公开化了。操作性定义使得概念从个人化的感觉和直觉中分离出来，并且允许任何实施可测量操作的人对概念进行检验。

例如，把“饥饿”这个概念定义为“我胃里不好受的感觉”，并不是一个操作性定义，因为它与个人对于“不好受的感觉”的体验相联系，因此不能被其他观察者知悉。相反，涉及一些可测量的食物剥夺时间或者像血糖水平这样的生理指标的定义才是操作性的，因为它包含了任何人都可以实施的可观测的测量。同样，心理学家不同意将“焦虑”定义为“我不时会感到的不舒服和紧张”，而是必须用像问卷和生理指标测量这样的操作来定义概念。上述那个定义是个人对身体状况的解释，他人无法复制；而后者则是把概念放在公共科学领域进行解释。

在科学领域里，定义一个概念靠的是一*系列*操作，而非单独的行为事件或任务，意识到这点非常重要。相反，一些差别细微的任务和行为事件通常聚合在一个概念上（在第 8 章我们将会更多地讨论聚合性操作）。例如，教育心理学家根据利用诸如“伍德库克阅读能力量表”之类（Woodcock，1998）的标准化工具测得的成绩来定义“*阅读能力*”这个概念。该量表测出的阅读能力总分包含了一些不同分量表测得的能力指标。这些分量表测查的能力稍有不同，但全都与阅读相关。例如，阅读一篇文章、想出一个合适的单词在文章中填空、写出一个词的同义词、独立拼读一个较难的词，等等。所有这些任务上的表现综合地定义了“*阅读能力*”这个概念。

操作性定义促使我们认真地、经验性地思考我们如何定义一个概念，所谓经验性，是指要根据我们对真实世界的观察。试想我们要给一个看起来相当简单的概念“打字能力”下一个操作性定义。想象一下你这么做是为了比较两种打字教学方法的优劣。思考一下你所要做的所有决定。当然，你想要测量“打字速度”。但是要打多长的一段文章呢？仅有 100 个单词的文章可能太短，而 10000 个单词的文章又似乎太长。那么到底多长才算好呢？打字速度维持多久才最符合我们对*打字能力*这一概念的理论建构？用什么类型的文章来测试呢？文章是否要包含数字、公式和不常见的间距？我们如何处理错误？当我们测量打字能力的时候，时间和错误似乎都应被考虑在内，但是，如果把这两个指标同时考虑进去的话，要如何来计算一个总分呢？我们想要让时间和错误具有相同的权重，

还是一个比另一个更重要？寻求一个好的操作性定义会迫使你认真考虑所有这一切；它会让你对如何将“打字能力”进行概念化做一番透彻的思考。

信度和效度

概念的操作性定义要想有用，必须同时具备信度和效度。信度是指测量工具的一致性。如果你对同一概念进行多次测评，是否能够得到相同的测量结果。信度的科学概念很容易理解，因为它与常识中的定义以及字典里的定义非常相似："任何总能够产生相同结果的系统所具备的一种属性"。

试想一下，一个外行人士会如何评价一件事是否可信呢？想象一个每天早上要赶公共汽车从新泽西去曼哈顿上班的人。按照时间表，公共汽车每天应该在上午 7:20 到达此人等车的站点。在一个星期中，如果公共汽车到达的时间分别是 7:20、7:21、7:20、7:19 和 7:20，那么我们就可以说在那一周汽车的到达时间是可信的，如果下周汽车到达的时间分别是 7:35、7:10、7:45、7:55 和 7:05，那么我们就可以说在那一周汽车的到达时间是非常不可信的。

在科学领域中，一个操作性定义的信度以类似的方式来评估。如果我们多次测量同一概念得到的结果是近似的，那么我们就说测量工具表现出较高的信度。如果在同一星期的周一、周三和周五，用同一 IQ 测验的不同版本测量同一个人的智力，得到的分数分别是 110、109、110，那么我们可以说这一 IQ 测试是非常有信度的。相反，如果三个测试分数分别是 89、130 和 105，那么我们就可以说这一 IQ 测试没有显示出高的信度。有一些专门的统计方法可以评估不同类型的测量工具的信度，所有标准的方法论入门教材中都有介绍。

但是请记住，信度仅仅是指前后一致，而不包括其他内容。对于一个操作性定义而言，仅有信度是不够的，信度是一个必要而非充分条件。作为一个好的操作性定义，操作必须被证明对于概念来说是有效的测量。

“结构效度”这个术语是指一个测量工具（操作性定义）是否测量了它本应测量的内容。保罗·考兹比教授（Cozby，2006）在其所著的方法论教材中为我们讲述了一个只有信度而没有效度的搞笑例子。假设你想测测自己的智力，测试者让你站到一个类似鞋码器的测试仪器上，然后仪器给出一个读数。当然，你会认为这是一个笑话。但是请注意，这个测量工具可以显示许多类型的信度，而这些信度在方法论教材中都会讨论到。这个仪器在星期一、星期三和星期五会呈现出相当一致的读数（这称之为“重测信度”），并且无论谁操作它，它都会给出一样的读数（称之为“评分者信度”）。

用鞋码器来测量智力，其问题不在于信度（这是有信度的），而在于效度。它不是一个测量它本应测量的概念（智力）的合理方式。断定它不是测量智力的有效方式的证据之一，就是我们发现它和其他一些被认为与智力相关的变量无关。鞋码器的测量结果与学业成就无关，与脑功能的神经生理学测量无关，与职场成功无关，与认知心理学家提出的信息加工效率的指标无关；相反，真正的智力测验与所有这一切都有关（Deary，2000；Geary，2005；Lubinski，2004）。在心理学领域，真正的智力测验要兼顾效度与信度，而智力的鞋码器测验只有信度而没有效度。

在这一点上，你可能想知道信度和效度的其他组合方式是否可行。因此，让我来重申一下我们的立场。在操作性定义中，我们寻求信度和效度兼备，因此高信度和高效度结合才是理想的目标。我们刚刚讨论了鞋码的 IQ 测试，目的是论证高信度和低效度是没用的。第三种情况是低信度和低效度，这绝对没有用，因此不值得讨论。但是你可能想知道第四种，也就是最后一种可能的组合方式：如果高效度和低信度又怎么样呢？答案是，和低效度和高信度的例子（鞋码器例子）一样，这种组合也是没用的。事实上，更准确的说法是，这类情况压根儿不可能出现。因为，如果不能进行可信的测量，你根本无法宣称测量是有效的。

直接和间接的操作性定义

概念和可观测的操作之间的联系，在直接和间接性程度上变化很大。很少有科学概念几乎完全是通过可观测的操作来定义的。大部分概念的定义采用更为间接的方式。例如，一些概念的使用既取决于一系列的操作，又取决于它和其他概念之间的特殊关系。最后，还有一些概念不通过可观测的操作直接定义，而是通过它与另外一些概念间的关系来定义的。这种概念有时被称为“潜在概念”，在心理学中非常普遍。

举个例子来说，许多研究关注所谓的 A 型行为模式，因为它与冠心病的发病率有关（Austin & Deary，2002；Curtis & O'Keefe，2002；Matthews，2005；Smith，2003；Suls & Bunde，2005）。在第 8 章中，我们将会更加详细地讨论 A 型行为模式。但是，这里重点要说的是，A 型行为模式实际是通过一系列二级概念来定义的：强烈的竞争欲望、潜在的敌意、赶时间行为、达成目标的强烈驱力等等。然而，每一个用于界定 A 型行为模式特征的概念本身也都需要操作性定义。事实上，研究者们已经为对每个概念进行操作性定义而付出了很多努力。我们讨论的要点是，A 型行为模式是一个复杂的概念，它并不是被操作所直接定义的。相反，该概念与其他一些各自具有操作性定义的概念联系在一起。A 型行为模式提供了一个间接操作性定义的例子。尽管不同的概念与可观测操作的联系程度各有不同，但所有的概念都在一定程度上通过其与可观测操作之间的联系来获得意义。

科学概念的演进

一个科学概念的定义并不是固定不变的，而是随着相关观测结果的不断丰富而发生变化。意识到这一点非常重要。如果一个概念的原始操作性定义在理论上被证明是无效的，那么该定义就会被抛弃，以另外一套定义的操作取而代之。这样，随着相关知识的积累，科学概念不断演进，其抽象性逐渐增加。例如，在一段时间里，人们认为电子是一个围绕原

子核旋转的带负电的微小球体。而如今，电子被视做在特定实验条件下，具有似波特性的概率密度函数。

在心理学领域，智力概念的发展提供了一个类似的例子。起先，智力仅有一个严格的操作性定义：智力是通过心理功能测验所测到的东西。随着实验证据的不断积累，智力被证明与学业成就、学习、脑损伤、神经生理学及其他行为和生物学变量有关，这一概念在逐渐丰富的同时又得到了提炼（Deary，2000，2001；Geary，2005；Lubinski，2004；Sternberg，2000；Sternberg & Grigorenko，2002；Sternberg & Kaufman，1998；Unsworth & Engle，2005）。现在看来，在定义智力概念时，最好用一种高等级的建构，通过多种更为具体的信息加工操作来定义。当然，这些假设的信息加工过程应该具备更为直接的操作性定义，可以用可测量的指标来表述。

人类记忆理论中的概念也以同样的方式发展。现代心理学家很少使用类似"记忆"或"遗忘"这样的笼统概念；相反，他们测量那些可以进行明确定义的记忆子过程，如短时听觉记忆、符号存储、语义记忆以及情景记忆。传统的"记忆"或"遗忘"的概念通过更加明确的操作性概念得到了细化。

因此，理论术语的用法在科学实践中不断演进，而不是在针对文字意义的争论中获得发展。这是科学的操作态度和本质主义者在追求绝对定义之间最显著的区别。神经病学家诺曼·格什温德（Geschwind，1985）将这种区别描述如下："我认为，从医学发展的历史中，你能够了解一件事，那就是：许多人认为，研究一个问题的方式就是首先定义这个问题然后再去研究它。这种想法被一次次地证明是错误的，因为你发现，只有先知道了答案，才能正确定义问题"（p. 15）。

哲学家保罗·邱吉兰德（Churchland，1988）强调，在科学中，对概念的定义不是源于文字界定，而是源于与之相关的观察和其他概念：

> 要想完全理解"电场"这个概念，我们就必须熟悉这一表述所处的理论原则体系，它们会共同告诉我们，电场是什么、做什么。

> 这是一个典型的例子。通常来讲，理论性术语的意义不是从单一的、具体描述其所适用的必要充分条件的定义中获得的，它们往往通过所在的理论原则体系而被间接地定义。(p. 56)

随着科学概念的演进，概念常常与许多不同的理论体系交织在一起，并且获得多种操作性定义。这种情况的出现并不是因为概念本身出了问题。例如，许多人认为心理学不可信，因为心理学中许多重要的理论概念——例如智力，可以用不止一种方法来操作化和概念化（Sternberg, 2000）。但这种情形并非心理学所独有，也不是一件令人绝望或束手无策的事情。事实上，在科学领域里，这种情况是普遍存在的。例如，“热”既可以从热力学理论，也可以从动力学理论的角度来概念化。物理学并未因此遭到贬斥。想想电子，它的许多特性都是以波的概念来解释的。可是，如果将其视为粒子，它的另外一些属性则更好理解。到目前为止，还没有一个人会因为物理学中存在着这些多重概念化现象就提出要抛弃它。

心理学领域的操作性定义

许多人在思考物理学或化学的时候，能够理解操作主义的必要性。他们知道，如果科学家准备谈论某一类型的化学反应、能量或者磁场，就必须有相应的方法来测量。不幸的是，当人们谈到心理学的时候，却经常无法认识到操作主义的必要性。为什么人们没有同样地认识到这一显而易见的事实：为了成为科学理论中有用的解释体系，心理学术语必须被直接或间接地操作化定义？

人们对心理学产生误解的原因之一，就是心理学上所说的“预设偏见”。在第 1 章中我们提到过这个问题。人们不会出于执着于某种关于岩石性质的信念来研究地质学，而在心理学中，情况就大为不同了。我们每个人都有关于人格和人类行为的直觉理论，我们用它们来“解释”

我们自己以及其他人的行为。我们所有的个人心理学理论里都包含着理论性概念（例如聪明、攻击和焦虑）。因此很自然人们会发问：为何我们必须接受一些其他的定义。尽管这种态度从表面上看来是合理的，但对于任何致力于理解人类行为的科学来说，它都是一个巨大障碍，也是公众对心理学产生困惑的一个原因。

误解产生的最主要原因，也即媒体在准确呈现心理学成果方面最大的障碍，这就是，心理学中的许多专业概念都是用日常用语来表达的。这些日常用语为大量误解的传播敞开了大门。外行人很少意识到，当心理学家把“智力”、“焦虑”、“攻击”、“依恋”等词语当作理论性概念来使用时，它们的含义和大众平常所说的意思不一定相同。

从之前关于操作主义的讨论中就能看出这种区别的本质。当在心理学理论中使用如“智力”、“焦虑”这些术语时，它们直接或间接的操作性定义决定了它们的正确用法。那些定义常常具有高度技术性，通常具有特定意义，并且在许多方面都不同于这些术语在日常生活中的运用。例如，当我们听到“对大样本的认知任务进行因素分析所得到的第一个主成分”这段话时，许多人都意识不到它是术语“智力”的操作性定义。

同样，如果是外行人使用术语“抑郁”，那么就意味着“感觉糟透了”。相反，在《精神疾病诊断与统计手册》（*Diagnostic and Statistical Manual of Mental Disorders*）中，对抑郁症的专门定义占用了超过 12 页的篇幅（American Psychiatric Association，1994），并且与“感觉糟透了”有着很大的区别。临床心理学家所谓的抑郁，并不等同于外行人所说的抑郁（Hollon，Thase，& Markowitz，2002）。在其他科学领域也都存在同样的问题，尽管可能没有心理学这么严重。回想一下前面对“生命”概念的讨论。正如米德瓦等（1983）指出的，问题在于，像科学中的其他专业术语一样，“生命”一词出自人们的日常用语，但在科学场合中的含义已远不同于日常谈话中的用法（p. 66）。

物理学家丽萨·兰道（Randall，2005）曾讨论过这类问题如何阻碍了公众对物理学的理解。她指出爱因斯坦相对论中的“相对性”一词被公众理解为“绝对是不存在的，因为任何事物都是相对的”，而事

实上，该理论的意思正好相反！兰道指出，实际上爱因斯坦的相对论认为：“尽管测量的实施者在测量时需要依赖他的参考物和参考系，但事实上他所测量的物理现象总有一个恒定的属性，这一属性超越了观察者的特定参考系”。爱因斯坦的相对论实际上是寻找物理现象的恒定属性的。事实上，爱因斯坦也认为他的这一理论如果被命名为“恒定论”会更加贴切。但是，“相对性”一词的地位在当时已经太过深入人心而难以改变了（p. 13）。

兰道继续指出，即使在物理学中，“模糊的词语选择也是造成某些误解的根源，科学家经常使用一些口语化的术语。他们为这些用语赋予特定意义，但没有经过正规训练的人是不可能想到这种意义的”（p. 13）。在心理学里也是如此。当心理学家和外行人用同一词语来表达不同含义的时候，他们之间常常产生误解。如果有新的词语产生出来用以描述心理概念，这样的困扰可能会少一些。有时会有这样的词语出现，正如物理学家有了“尔格”和“焦耳”一样，心理学家有了“失调”和“编码”，这些词不是凭空编造的，但在日常用语中比较生僻，从而可以防止混淆。

“但是，”外行人可能提出这样的反对，“为什么心理学家这样折磨我们？为什么有这么多新的术语、高度专业性的定义、生僻的词语？为什么我们需要这些？为什么我对‘智力’概念的定义得不到他们的认可呢？”

在这里，我们来看一个对心理学研究有严重误解的例子——这一误解经常出现在关于心理学研究的媒体报道中。一份全国性的报纸以“你能用一般人听得懂的话重新说一遍吗？”为标题报道了1996年美国心理学会的一次会议，并说“心理学家所用的语言只有他们自己能听懂”。该文嘲讽了在会上报告的一份题为《用Gf-Gc理论解释对WJ-R和KAIT的联合因素分析》的论文。尽管记者表示他“不敢贸然猜测这个标题的真正意思”，但几乎所有接受过培训的心理学家都能理解这个标题是有关智力测验理论方面新进展的。的确如此。Gf-Gc理论是智力理论方面的一个技术性进展，记者没有理由听到过这个概念——就如同

我们不会期望该记者知道物理学家最近刚发现的一种基本粒子的细节一样。可是，有时候，记者对科学术语的无知（这是完全可以理解的）却对现代心理学造成了负面影响。

我们来看看问题的症结所在。解决它的第一步，就是强调我们已经讨论过的一个观点：操作主义不是心理学所独有的，它是所有科学门类的特征。大多数情况下，我们很容易接受它，理解它的显而易见的本质。如果一个科学家是研究放射性的，我们会理所当然地认为他肯定有办法测量此种现象——其他研究者也能使用该方法获得相同的结果。操作定义因此导致科学的公开化，而公开化是科学的关键特征之一。如果两个科学家对同一个操作性定义达成一致，其中一个人就可以用它去复制另一个人的结果。但是，在其他情况下看来显而易见的事情，在我们谈到心理学的时候却不怎么明晰了。人们经常意识不到“智力”和“焦虑”这些概念的操作性定义的必要性，因为我们总是在使用这些术语，难道我们还不“知道”它们是什么意思吗？

答案是：“是的，我们确实不知道它们是什么意思”——不是从科学家必须知道的意义上，而是从公众的意义上。一个科学家必须通过如下方式“知道”智力是什么意思：他必须精确地定义一种方法，使其他实验者能够以完全相同的方法测量这一概念，并且得到有关此概念的相同结论。就其明确性和精确性来说，这与日常交谈中为了实现相互理解而使用的模糊语言间有很大差别。

作为人性化力量的操作主义

如果过分依赖于我们“知道”的东西，肯定会产生问题，这个问题同样困扰着所有的直觉（非经验）信仰体系。关于某个事物你所“知道”的和张三、李四所知道的可能并不一样，我们如何决定谁是正确的呢？你或许会说“我强烈地感觉到我所知道的是正确的”，但是，如果张三的观点和你有出入，但比你拥有更强烈的感受呢？李四的观点与你俩都不同，也宣称自己是正确的，因为他的感受甚至比张三还要强烈。

讲这个简单的小段子，仅仅是想阐述科学知识的一个基本特点，它在人类历史中是一股重要的人性化力量：在科学中，知识的正确与否并不取决于个体提出主张时自己的肯定程度。所有建立在“直觉”基础上的信念体系都有一个共同的问题，即当出现矛盾观点时，它们缺乏一种机制来判别哪个是对的，哪个是错的。因为人人都凭直觉认为自己是对的，但当大家的直觉观点发生冲突时，我们该如何决定谁正确呢？令人悲哀的是，历史表明，这种冲突的结果通常是权力斗争。

一些人错误地宣称，心理学的操作取向使人们丧失了人性，而且我们应该把我们关于人类的观点建立在直觉基础之上。心理学家唐纳德·布罗德本特（Donald Broadbent）在 1973 年论证说，真正人性化的观点是将关于人类的理论观点建立在可观测的行为基础上，而不是以理论者的直觉为基础：

> 除非我亲眼看到别人在特定情况下亲自做了或说了什么，否则无法对别人做出判断……实证主义的方法是一种调和差异的方式。如果拒绝这一方式，那么处理争论的唯一方式就是面红耳赤的争辩了。（p. 206）

因此，科学中人性化的力量就是让知识公开化，让任何有冲突的观点都能以一种双方都接受的方式得以检验。回想第 1 章中提到的“重复”的概念。这让我们可以通过一种大家都事先同意的、平和的方式来从理论中进行选择。科学的公共性本质在很大程度上依赖于操作主义的理念。通过对概念操作化的界定，概念进入了公共的领域——任何人都可以对其进行批判、验证、改进或否定。

心理学概念不能以某些人的个人定义为基础，因为这类定义可能是不常见的、个人化的或者模糊的。由于这个原因，心理学必须摒弃所有对概念所做出的个人化定义（就像物理学拒绝对能量的个人化定义，气象学拒绝对云的个人化定义），而坚持公众可以知悉的定义，这种定义用操作来界定概念，并确保了任何一个接受过适当训练并拥有适当设备的人都可以实施这些操作。就摒弃个人化定义而言，心理学并没有将外

行人拒之门外，而是将这一领域向公众敞开，就像所有学科那样，以期寻求所有人都可以共享的、普遍的、公众可以利用的知识。

只有当概念以操作性定义为基础，并且不关注于本质主义者所讨论的文字意义时，这类具有公众可用性的知识才能够用来解决人类的问题。例如，蒙克（Monk，1990）描述了二战期间“创伤性休克”这个概念是如何在医学领域产生问题的。一些医师对此症状的诊断依据是过高的血红细胞浓度，并认为其原因在于血液中的血浆渗透到了组织中。其他医师诊断“创伤性休克”则根据低血压、皮肤苍白和脉搏过速。换言之，医生们对这一概念的操作性定义是不一致的（甚至是带有个人色彩的），因此，英国医学研究会的格兰特（Grant）医生建议说，“创伤性休克”这个概念应该被抛弃，并且对伤者进行详细观察时也不使用这个术语……由于在诊断方面缺乏共同的基础，无法对各种治疗措施的效果进行评估”（Monk，1990，pp. 445-446）。换句话说，这种概念弊大于利，因为缺乏一个获得普遍认同的定义使之成为公共知识（也就是被广泛地分享与认同）。

有时候，在科学领域中，概念意义的改变会导致对这一概念的科学理解与外行人士的理解产生冲突。法伯和邱吉兰德（Farber & Churchland，1995）讨论过一个关于“火”这一概念的情况。经典的概念是这样定义火的：“不仅是含碳物质的燃烧，而且还包含了太阳及各种星体上的活动（实际是核聚变）、闪电（实际上是电引起的白热化现象）、北极光（实际是光谱发射）、萤火虫的闪光（实际上是发出磷光）。在现代概念体系中，这些现象都与氧化无关，因此和木材燃烧不属于同一类型。另外，有一些现象原本认为是与燃烧没有任何关系（由于那时放热被认为是燃烧的本质特征）的过程——如生锈、锈蚀和新陈代谢——却被证实属于氧化现象”（p. 1296）。总之，氧化的原则使得篝火和生锈联系了起来，而将闪电与它们区分开来。对于科学家而言，这也许是一个进步的标志，但却让外行人士感到迷惑和无所适从了。

本质主义问题和对心理学的误解

许多人在接触心理学时放弃操作主义观点的另一个原因是，他们想为这些问题找出本质主义的答案。这样做到底是因为心理学新近才从哲学中分离出来，还是因为公众对心理学了解得比其他学科少，尚不得而知。然而从某种意义上讲，这并不重要。最终的结果是一样的。人们期望心理学可以就其他学科所不能回答的这些复杂问题给出终极答案。

回想本章开头提出的问题："地心引力"这个术语的真正含义是什么？它的内在本质是什么？在谈到地心引力一词时，我们到底指的是什么呢？大部分人认为这些问题需要绝对性的知识，需要理解现象的潜在本质，而物理学当前的理论不能对这类问题提供答案。对关于物理科学近几百年来发展的通俗读物比较熟悉的人都能意识到，地心引力是一个高度复杂的理论建构，并且其概念性和操作性关系也处在不断变化之中。

可是，如果将上述问题中的"地心引力"全都换成"智力"，奇迹就出现了。现在那些问题立刻被赋予了重大意义。它们看起来是那么自然和富有深意，它们就是在寻求一个终极答案。可是当心理学家给出和物理学家一样的答案，即"智力是一个复杂的概念，它的意义是由测量它的操作以及它与其他概念之间的理论关系来界定的"时，却会被鄙视和指责为回避真实问题。

心理学所面临的一个难题就是，公众要求心理学去回答本质主义问题，而通常其他科学家并不需要回答类似的问题。这类要求常常导致人们贬低心理学领域已经取得的进步。尽管这类要求不能阻止这一领域自身的发展——因为就像其他科学家一样，心理学家无视本质主义问题并继续他们的工作——但那些问题成了公众理解心理学的障碍。当一个不了解情况的批评家声称心理学没有取得进步时，公众就会迷惑了。这类责难极少遇到挑战，这也反映了本书序言中所述的不幸事实：对于心理学领域所取得的科学成就及其意义，公众的了解是极度匮乏的。当我们仔细审视那些对心理学的批评，不难发现它们通常归结于一点：心理学至今没有为它提出的问题提供终极答案。对于这种指责，心理学毫不犹

豫地低头认罪——像所有其他科学学科一样。

一些人可能会很不舒服地发现，包括心理学在内，没有任何科学可以对本质主义问题做出回答。霍尔顿和罗勒（Holton & Roller，1958）讨论过，当外行人被告知物理学不能够回答本质主义问题时所表现出的那种不安。他们谈论的是与放射性衰变有关的现象：发生衰变的放射性元素的原子数量与时间是呈指数函数关系的。可是，这种函数并不能解释为什么放射性衰变现象会发生。这个问题的解决将可能再次涉及另一个数学函数，但是它还是不能告诉外行人什么是真正的放射性衰变。霍尔顿和罗勒告诉我们："我们必须平静地接受现代科学的局限性，它并没有声称可以发现'事物究竟是什么'"（pp. 219-220）。科学作家罗伯特·怀特（Wright，1988）解释说：

> *伊萨克·牛顿的地心引力理论有些地方不尽人意……毕竟，"在一定距离外作用"如何实现？牛顿回避了这样的问题……自从牛顿开始，物理学家们一直在仿效他的做法……物理学家们不再尝试解释为什么事物必须遵守电磁学规律或地心引力规律。*（p. 61）

同样，如果那些为人类本性问题寻求本质主义答案的人求诸于心理学，注定将会失望。心理学不是宗教，它是一个试图对所有行为做出科学解释的广阔领域。因此，心理学现在的解释是对行为的暂时性的理论建构，就目前来说，这些建构在解释行为方面优于其他解释。这些建构在将来注定会被更好的、更接近事实的理论概念体系所取代。

操作主义和心理学问题的措辞

在评估一个心理学理论的可证伪性时，操作性定义的理念是一个非常有用的工具。概念有没有直接或间接地建立在可观测操作的基础上，是识别不可证伪的理论的重要线索。没有建立在可观测操作基础之上的概念，通常是为了拯救那些不能被数据印证的理论。所以，那些不严格的概念——理论学家不能为它们提供直接或间接的操作性联系——都应

该引起怀疑。

与之相关的是科学家称之为“节省”的原则。节省原则是指，当两个理论有同样的解释能力的时候，较为简单的理论（涉及更少的概念和概念性关联）胜出。原因是，拥有较少概念性关联的理论在将来的检验中会更具可证伪性。

深刻理解操作主义的原则，也有助于我们识别不具备科学意义的问题。例如，在我的电脑文件夹里，有一篇来自于国际联合出版社的在线服务文章，标题为《动物会思考吗？》。这篇文章讲述了动物行为方面最新的实验。文章中所引述的研究没有任何错误，但是，显然这个标题仅仅是一个玩笑。这个标题的问题在于没有科学意义，没有关于“思考”的操作性标准。许多报纸的标题中存在类似的问题，比如“计算机会思考吗？”没有操作性标准的话，这个问题也没有科学意义，但在鸡尾酒会上倒是可以大派用场。

实际上，观察人们在后面这个问题上的争论具有启发意义，因为这样的一个争论为我们亲眼见证先前讨论过的心理学中的“预设偏见”问题提供了机会。大部分人都有一个强烈的预设偏见，不希望计算机能够思考，为什么呢？出于各种原因，外行人认为“思考”这个概念与“人类”这个概念紧密联系，许多人在情感上不能接受非人类的物体（例如，计算机或看起来不像地球人类的外星生命形式）也能思考。

可是，尽管大部分人对会思考的计算机这一设想表示强烈反对，但他们并没有认真思考这一问题，也没有对“思考”做出一个更好的定义，使其能包含大部分人类（例如，婴儿）而排除所有计算机。有时，那些不熟悉人工智能进展的人提出的定义听起来颇为滑稽，因为他们总是选择了计算机能够做的事情作为标准。例如，许多人提出以“从经验中学习的能力”为标准，但一些计算机和人工智能已经达到这个标准了（Churchland，1995；Clark，2001；McCorduck，2004；Pfeifer & Scheier，1999）。预设偏见的力量在这种情况下显而易见。会有人这样反应吗？“哦，我不知道。既然有些计算机符合我提出的‘思考’的标准，那么我不得不说至少有些计算机是能够思考的吧！”通常大家是不会做

出这样理智而诚实的反应的。更常见的情况是，人们开始寻找另一种标准，并期望计算机不能满足该标准。

通常人们的第二选择是“创造性”（“想出一些有用而且没有人想到过的东西”——我们先不管大部分人是否满足这个标准）。当被告知大多数专家都认可计算机能够达到这个标准时（Boden，2003；Pfeifer & Scheier，1999），人们仍然不愿承认机器思考的可能性。人们常常不会想到要做出一个操作性定义，转而提出计算机是不可能思考的，因为“是人类制造了计算机并且设计了程序；计算机只是执行程序而已”。

尽管这是反对机器思考的最古老的观点之一（McCorduck，2004；Robinson，1992；Woolley，2000），但它还是错误的。预设偏见让许多人认识不到，这些辩解与要讨论的问题毫不相干。几乎每一个人都认为思考是发生在自然世界中的一个过程。现在请注意，我们在讨论其他过程时，并不涉及“起源”的问题。考虑一下食物加热的过程。想想这个问题：“炉子能加热吗？”我们会说：“炉子不能加热，因为炉子是被人类制造出来的。因此只能说是‘人’在加热。真正加热的不是炉子。”或者，什么是“举重”？起重机能“举重”吗？我们的答案是否仍旧为“起重机不会举重，因为起重机是由人造出来的。因此，我们只能说是‘人’在举重。起重机真的不能举重”？当然不能这样说。一些事物的起源与它执行某一特定任务的能力是完全不相关的。思考的过程也是如此。一个事物能否思考并不依赖于这个事物的起源。

因为人们无法理性地接受机器有思考的能力，著名的计算机科学家艾伦·图灵（Alan Turing）设计出著名的“计算机能否思考”的实验。图灵设计的实验是一个操作性的实验，这对于我们的讨论是非常重要的。1950 年，图灵在题为《计算机器与智能》的著名文章中写道，“我建议去考虑‘机器可以思考吗？’这个问题。”他不想在鸡尾酒会那样的场合随意谈论这个问题，也不想如本质主义者那样无休止地讨论“思考”是什么意思，而是提出一个严格的操作性测试。他的想法是：如果计算机能够进行智能对话的话，那么就可以说它是有思考能力的。

图灵这一设想中的创造性在于，他提出一种方式将问题变得可操作

化，同时又防止了“预设偏见”的干扰。图灵对于检验计算机是否可以进行智能对话的测试逻辑进行了严格的限定。这个测试并不是让测试者通过键盘和屏幕与计算机互动，然后由测试者判断计算机有没有进行智能对话。图灵没有采用这种设计，因为他很担心“预设偏见”的干扰。图灵确信，一旦一个人坐到计算机、键盘和屏幕（显然是一些机器）之前，无论这个机器做什么，这个人都会否认它有思考能力。因此，图灵提出应控制与思考能力无关的外在因素。其著名的实验设计是让测试者通过两个键盘对话（一个和计算机相连，另一个和人相连，并且都在视线之外），然后再判断哪个是人，哪个是机器。如果被试不能以大于随机水平的正确率猜出哪一个是人，那么我们就有理由推断计算机具有同人一样的对话能力，而对话能力正是“思考”的操作性定义。

图灵的主要思路“与交响乐团选拔乐师的试听面试的思路是一样的，试听面试时，在评委与面试乐师之间放置一个不透明的屏幕，前者要隔着这个屏幕来判断乐师演奏的好坏。很显然，评委关心的是音乐能力，而且也仅仅是音乐能力而已。性别、头发的长度、皮肤颜色和体重等都是完全不相关的……图灵认为，人们对智力的判断可能同样受到对方是否拥有柔软的皮肤、温热的血液、面部特征、手和眼睛等那些明显不是智力本质的因素的影响”。（Dennett，1998，p. 5）图灵的测试启发我们，如果我们想要理性地讨论心理学概念，那么操作性定义是必需的；我们要以一种有条理的方式进行判断，而不仅仅根据我们自己对某个问题的偏见。

观察人们讨论人工智能问题时所展示的思维方式，就能发现科学和非科学思维方式之间的区别。科学的方式是先发展一种合理的操作性定义，然后看我们可以从中得出哪些关于思考、计算机和人类的结论。与之不同的是，预设偏见主导了大多数人的思维。他们已经得出了某一结论，并且对于计算机和人类的表现之间已被发现的差异并不感兴趣。相反，一旦形成定势，他们就会绞尽脑汁地去找出各种理由，来巩固自己的这些想法，避免发生改变。于是，我们看到，正是预设偏见和非操作性的本质主义态度，让人们认定他们“就是知道”思维到底是怎么回事。

这种态度使大多数人的直觉心理理论无法证伪，因此完全无用。也正是这种态度说明了为什么我们需要科学心理学！

小　结

操作性定义是利用可测量、可观察的操作来表述的概念定义。我们确信某个理论具有可证伪性的主要途径之一，就是确定理论中的关键概念具备可用可重复性很强的行为观察来表述的操作性定义。操作性定义是让科学知识变得公开可检验的主要机制。这样的定义被置于公共领域，使其所界定的理论性概念能够接受所有人的检验，而不是像“直觉的”、非经验性的定义那样，只属于特定个体，检验它的机会并不向所有人开放。

由于心理学使用一些来源于日常生活的词语，如智力和焦虑，许多人对于这些术语的含义有着预设的想法，因此往往意识不到对这些术语进行操作性定义的必要性。心理学和所有其他科学门类一样，也需要对其术语进行操作性定义。可是，人们常常要求心理学家回答本质主义的问题（有关概念的纯粹深层本质的问题），而其他科学家就不必回答这类问题。没有科学能够回答这样的终极问题。心理学和其他科学门类一样，正在试图不断地完善其操作性定义，使理论概念能够更加准确地反映真实世界的原貌。

4

见证和个案研究证据

——安慰剂效应和了不起的兰迪

画面切换到奥普拉秀——过去十年中最著名的电视脱口秀节目之一——的现场。今天的嘉宾是俄狄浦斯人类潜能研究所的所长阿尔弗雷德·庞蒂菲科特（Alfred Pontificate）博士 。这位博士新提出了一个有关出生次序的激进理论，这一理论的基本理念是：个体的生命进程是被家庭互动所设定的，而家庭互动是由出生次序决定的。奥普拉鼓励观众对此理论进行提问。讨论无可避免地由最初的理论关注，转向了为观众个人生活中的重要事件做出解释。这位博士欣然应允。

例如，"博士，我的哥哥是个不要命的工作狂。他对妻子和家庭完全不管不顾，并且把与工作有关的问题看得比什么都重。他有溃疡和酗酒问题，但他拒不承认。他们家在近两年内从没过过一个真正意义上的假期。他的婚姻也快玩完了，但他似乎并不是特别在乎。他为什么要选择这样一种自我毁灭式的生活呢？"

博士反问道："亲爱的，他在家中排行第几？"

"哦，他是子女中的老大。"

"这就对了，"博士说道，"这在生活中比较常见。我们在临床上经常见到这种现象。这类现象出现的深层次原因是，父母将自身的愿望和

挫折都转移到他们第一个出生的孩子身上。通过愿望的这种无意识的转化过程，即使父母从未明确要求过孩子，孩子也在内化这些愿望和挫折。然后，通过这种我称之为‘期望上旋’的动力过程，父母的抱负转化为孩子对于成功的病态的渴求。”

当嘉宾挑战观众的信念时，奥普拉秀的观众有时会提一些尖锐的问题，但当行为“专家”似乎是在印证观众的传统观念的时候，这种情况就很少发生。然而曾经有过那么一次，节目因为一位观众质疑嘉宾的主张而显得异彩纷呈。有一位热切而直率的观众当时正身处演播室，“但是请等一下，博士，”提问者开始了他的问题，“我的哥哥也是家里的老大。我的父母把那个笨蛋送到哈佛，而让我去了一个将来能够成为一名牙医的两年制专科学校。但他们的‘神童’在一年之后就辍学了，跑到了科罗拉多州的山顶上。我们最后一次见到他时，他正在编篮子！我搞不懂你关于‘长子’的说法。”

这位观众使现场气氛骤然紧张，但是博士总是能够逢凶化吉：“哦，是的，我也曾经见过很多像你哥哥一样的个案。是的，我经常可以在我的从业中遇到这样的人。他们的‘期望上旋’的动力过程发生中断，生成潜意识的要求来抵抗父母转化到他们身上的期望。这样的话，个体的生活规划会朝着与传统成就标准相反的方向发展。”一阵肃然的沉默之后，讨论转向了下一个“案例”。

这些场面我们都再熟悉不过了，只不过又是一个关于本杰明·拉什问题的例子罢了。关于出生次序的“理论”是在没有一个事例能够证明其“不成立”的思维框架下被构想出来的。由于它是一个不能证伪的理论，搬出再多能证明它的证据也没有意义，因为这个理论不能排除任何可能的情况。

然而，我们在本章所关注的并非这一理论本身，而在于那些用于支持它的证据。当被迫出示证据时，庞蒂菲科特博士搬出了他的“临床经验”或“个案研究”。这在媒体心理学领域是一个惯用的套路。脱口秀节目和通俗心理学图书中充斥着基于作者临床经验的心理学理论。他们通过这类渠道提供给公众的许多疗法，能够支持这些疗法的，无非是

那些曾接受治疗并认为得到了改善或被治愈了的人的个人见证。在本章中，我们将为心理学信息的消费者建立一个非常有用的原则：个案研究和见证作为评估心理学理论和治疗的证据是毫无价值的。

在本章中，我们将要证明这个原则为什么是正确的，并且还要讨论个案研究在心理学中的正确作用。

个案研究的地位

个案研究的作用，很大程度上取决于科学探索在某个特定领域进展到什么程度。从个案研究或临床经验中获得的灵感，在特定问题的早期研究阶段或许比较有用，因为它们可以提示哪些变量需要进一步研究。个案研究在开启心理学新的研究领域方面起到过关键作用。让·皮亚杰（Jean Piaget）的工作中就是很著名的例子。皮亚杰的研究提出了一种可能性，即儿童的思维并不只是成人思维的简易版或低级版，而是有其自身结构的。皮亚杰关于儿童思维的部分推测已经被证实，但很多还有待证实（Bjorklund，2004；Goswami，2004）。然而，对于我们这里的讨论来说，更重要的不是皮亚杰的哪些思想被证实了，而是要理解，皮亚杰的个案研究尽管没有证实任何事情，但它为发展心理学家的研究提供了难以置信的广阔领域。第 5 章和第 6 章中所要介绍的相关研究和实验研究，为皮亚杰个案研究中提出的假设提供了或支持或否定的证据。

然而，当我们从科学研究的早期阶段（在此阶段个案研究可能是极为有用的）步入更为成熟的理论检验阶段之后，情况就大大不同了。由于个案研究在特定理论的检验中不能作为证实或证伪的证据，所以它在科学研究的后期不再有效。其原因就是：个案研究和见证叙述都是所谓的“孤立事件”，缺乏比较性信息，而这种信息对于排除其他可能的解释来说是必要的。

见证叙述与个案研究相似，因为它们都是孤立事件。依赖见证叙述的问题在于，如果累积起来的见证能够为几乎每一种疗法提供支持，那

么它也就不可能用来支持任何一种特定的疗法，因为所有相互对立的疗法都有各自的见证。当然，我们想知道的是哪种疗法是最好的，但我们不能依据见证来决定。正如心理学家雷·尼克尔森（Ray Nickerson，1998）在评论我们用以欺骗自己的认知过程时所说的那样，"江湖郎中的骗术往往得逞，是因为他们总能找到一些病人愿意为他们做见证，这些病人总是发自内心地告诉别人，他们自己的确从治疗中获益匪浅"（p. 192）。例如，有大量的见证声称潜意识自助式录音带（用一种低于听觉阈限的信号制作出来的录音带）可以提高人的记忆力，甚至提高人的自尊，然而，在严格控制条件下进行的研究显示，这类录音带对记忆力或自尊没有任何改进（Greenwald，Spangenberg，Pratkanis，& Eskenazi，1991；Moore，1995）。

"其他可能的解释"这一理念，对于理解理论检验来说至关重要。实验设计的目标就是构建某一事件或现象，使其只能用某一种特定的理论来解释，而其他理论则解释不通。正如第 2 章所说的，只有当我们收集的数据排除了其他可能的解释时，科学才能进步。科学为理论观点的自然选择创设了条件。有些理论观点经过实证检验存活了下来，而另一些则被淘汰出局，凡保留下来的都更接近真理。但是，这是个慢工出细活的过程，各种理论观点都必须经过细致审查，以便发现哪些更接近真理。但是这一过程必须有所取舍：为支持某一特定理论所收集的数据，不能同时支持许多其他可能的解释。基于这一理由，科学家在他们的实验中设有控制组，或称为对照组，以期得到比较性信息。这样做的目的，是为了能够在比较控制组与实验组的结果时，排除其他可能的解释。至于实验设计如何能做到这一点，将是后面几章的主题。

个案研究和见证叙述作为孤立的现象而存在，它们缺少必要的比较性信息，不能证明某一特定的理论或疗法更优越。因此，引用某个见证叙述或个案研究的结果来支持某一特定理论或疗法是错误的。如果这么做的那些人不指明他们所提供的所谓证据其实也适用于大量其他可能的解释，那他们就是在误导公众。简言之，针对某个现象的孤证具有高度的误导性。安慰剂效应的例子将更清晰具体地阐释这一论点。

为什么见证叙述毫无价值：安慰剂效应

几乎每种产生于医学和心理学的疗法都有一定数量的支持者，并且总能催生出一些发自内心认可其疗效的人。医学文献记载了猪牙齿、鳄鱼粪便、埃及木乃伊的粉末，以及很多更富想象力的东西都曾经具有疗效（Harrington，1997；Shapiro，1960）。事实上，人们早已熟知，仅仅暗示正在接受某种治疗，就足以使许多人感觉病情好转了。

无论治疗是否有效，人们都会报告某种疗法曾经对他们有所帮助，这种倾向被称为*安慰剂效应*（Christensen，2001；Ernst & Abbot，1999；Harrington，1997；Kirsch，1999；Russo，2002；Stewart-Williams & Podd，2004）。安慰剂效应的概念在电影《绿野仙踪》中有绝佳的阐述。仙女并没有真的给铁皮人一个心脏，没有给稻草人一个大脑，也没有给狮子以勇气，但是他们都感觉更好了。实际上，直到近一百多年，医学才发展出较多具有确凿疗效证据的治疗方法，因此有人曾经这样说：“本世纪以前，整个医学史只能说是安慰剂效应的历史罢了”（Postman，1988，p. 96）。

我们可以通过对生物医学研究的考察来说明安慰剂效应这一概念。在生物医学研究中，所有的新药研究程序都必须包括对安慰剂效应的控制。一般来说，如果在一组病人身上试验一种新药，就要组建一个患同样病症的对等组，给他们服用等量不含任何药物的药剂（安慰剂）。两组病人都不知道他们吃的是什么药。这样，两组进行比较时，安慰剂效应——即给予病人任何一种新的治疗都会使他们感觉好些的这种倾向——就能得到控制了。仅仅报告百分之几的病人吃了新药后症状得以缓解是不够的，因为如果没有控制组的数据，就不知道报告症状缓解的病人是由于安慰剂效应还是药物本身的疗效。

安慰剂效应在抑郁症治疗中是29%（即29%的病人服用安慰剂后报告症状缓解了），在十二指肠溃疡中是36%，在偏头痛中是29%，食道炎是27%（Cho，Hotopf，& Wessely，2005）。一项研究（Bower，

1996b）建议，将安慰剂效应与当下流行的抗抑郁剂百忧解结合起来，将会最大限度地发挥药物自身的作用。安慰剂效应的效力是很强的，以至于曾有报告说有人对安慰剂成瘾（Bok，1974；Ernst & Abbot，1999），这些人需要服用剂量越来越大的安慰剂来保持他们的健康状态！

当然，在有关药物治疗的实际研究中，安慰剂控制并不只是一个什么都不含的药片，而是含有当前认为最有效的药用成分。实验比较的目的在于揭示，新药是不是比当前最有效的药还要好。

你每次吃处方药时都会得到安慰剂效应的提示信息，下次吃处方药的时候（如果你非常健康，就看看你祖母的药吧！），仔细查看一下药物附带的说明书（或者登陆药品制造商的网站浏览一下），你将在药物问题说明里看到安慰剂效应的信息。例如，我吃一种叫做 Imitrex（琥珀盐酸）的药物来缓解偏头痛。此药附带的说明书告诉我：控制研究已经证实，在服用一定量的药物之后，57% 的病人在两个小时之内其症状得到了缓解（我就是这幸运的 57% 之一！）。但是说明书同时告诉我，同样的研究显示，这类偏头痛中安慰剂效应是 21%——有 21% 的人在服药后两小时内症状得到缓解，即使他们服用的药物里是中性材料而非琥珀盐酸。

安慰剂效应在心理治疗中也很常见（Wampold et al.，1997）。许多有轻度和中度心理问题的人，在接受心理治疗后说他们的情况有所好转。然而控制研究证明：这一康复比例中，有相当一部分是由于安慰剂效应和时间推移这两个因素共同作用的结果，时间推移通常被称为*自然康复现象*。大多数有效的治疗都是由于治疗效果和安慰剂效应以某种不为人知的组合而产生的效果。正如多兹（Dodes，1997）指出的："即使严重的疾病也有恶化和缓解的时候；关节炎和多发性硬化症就是典型的例子。甚至癌症也会莫名其妙地消失"（p. 45）。他同时也警告说，对于安慰剂的积极反应并不意味着病人的病是虚构出来的，他还警告，与流行的观念正相反，安慰剂可以是有害的："安慰剂效应能够通过证实或强化想象中的疾病来'诱发'慢性病。病人会对那些利用安慰剂效应的非科学从业者产生依赖。"（Dodes，1997，p. 45）

在关于心理治疗效果的研究中，怎样合理地对待安慰剂效应控制组，往往令人颇费周折。但是，这些复杂的问题不是我们在这里所要关注的，理解研究者为什么要将药物治疗的真实效果与安慰剂效应及自然康复区分开却很重要。高登·保罗（Paul，1966，1967）关于治疗效果的研究为我们提供了例子，告诉我们这类研究结果揭示了什么。保罗调查了几组学生，他们“在公开场合说话时会产生不适和焦虑”。实验组接受了针对语言紧张问题的脱敏疗法，85% 的被试表现出显著的改善。安慰剂组拿到了一些药片，他们被告知这些药片是有效的镇定剂，但实际只是一个碳酸氢钠胶囊。在该组中，有 50% 的人表现出明显的好转。第三组根本没有接受任何的治疗，仍然有 22% 的人表现出明显的好转。这样看来，对于这一特定问题来说，自然康复比例为 22%，另外 28% 表现出的改善产生于安慰剂效应（50% 减去 22 %），脱敏疗法所具有的真实疗效则高于安慰剂和自然康复加起来的效果（85 % > 50%）。

和保罗的研究类似，其他研究也显示，心理疗法确实优于只用安慰剂所产生的效果（Hollon et al.，2002；Lipsey & Wilson，1993；Nathan & Gorman，1998；Shadish & Baldwin，2005）。但是，使用了安慰剂控制组的实验也表明，仅报告有多大比例的人感觉自己有所好转，会严重高估治疗的实际效果。问题就在于，得到见证叙述简直不费吹灰之力。康奈尔大学的心理学家托马斯·吉洛维奇（Thomas Gilovich，1991）指出：“人类拥有如此容易自愈的身体，即使医生不做任何事情，很多寻求医学帮助的人也将体验到积极的疗效。如此一来，当自然康复的比率很高的时候，即使是毫无价值的治疗手段也能显得有效”（p. 128）。简言之，无论干预的效果如何，只要运用治疗干预，潜在的安慰剂效应就会显现。问题在于，安慰剂效应是如此强大，以至于无论某个人使用的疗法多么荒唐，只要是被应用于一大群人的话，总有一些人会乐于为它的效果做出见证（清晨头部击打疗法，每天使用让你神清气爽！给我寄 10.95 美元，你就可以得到这个特制的、经过医学测试的橡胶锤）。

但我们确实不应该拿这种严肃的事情开玩笑。轻信见证叙述和个案研究的证据可能会导致灾难性的后果。曾为抽动性秽语症做出科学界定——将之定义为器质性紊乱——的研究小组（Shapiro et al.，1978；见第 2 章）指出，人们对于个案研究证据的错误依赖，使得关于该病的、不可被证伪的精神分析理论长期盘踞不去，阻碍了对于该病病理进行真正的科学研究。英国科学期刊《自然》（*Nature*）的华盛顿记者史蒂芬·巴蒂安斯基（Budiansky，1984）在总结医学领域中的这种现象时说了如下一段话，强调了很多我们在前面章节中提到过的一些科学的观点：

> 科学回避个人化的东西。虽然这种倾向常常被归结为科学家所应具备的冷静，但事实上，它确实是 20 世纪人类智慧的伟大战利品之一。科学家们深知，要探寻自然界的原因和结果，就必须排除个人的感觉及经验。健康科学曾经有过一个粗糙的阶段，人们生病有许多原因，大多数情况下，人们无论接受何种"治疗"之后，都会有所好转。各类庸医庸术之所以能够大行其道，凭借的就是那些对疗效满意的患者的见证叙述，这些活生生的例子，说明人们要跳出个人经验的束缚有多么困难。（p. 7）

发表在《新英格兰医学杂志》（*New England Journal of Medicine*）上的一篇社论，论述了在医学科学的从业者眼中个案研究和见证叙述的地位。"例如，如果这本杂志收到一篇论文，说一个患胰腺癌的病人在服用了大黄根（rhubarb）后康复了……我们可能会发表一篇个案报告，但是，我们发表它并不是宣告它为一种新的疗法，而仅仅是推荐它作为一个值得用正规的临床实验进行验证的假设。与之相反，关于各类偏方秘方的轶闻（通常发表在通俗书籍和杂志上）则没有做出这样的声明，并且这些轶闻本身也不足以作为支持那些疗效的文献。"（Angell & Kassirer，1998，pp. 839–840）。

“鲜活性”问题

安慰剂效应的存在，宣告了见证叙述作为证据是无效的，这么做尽管很痛快，但是我们必须意识到，还存在着另外一个障碍，它阻碍了人们理解这一问题。社会和认知心理学家已经研究了人类记忆和决策中的所谓“鲜活性效应”（Kunda，1999；Ruscio，2000；Sinaceur，Heath，& Cole，2005）。当面临问题解决或决策情境的时候，人们会从记忆中提取与当前情境有关的信息。因此，人们倾向于利用更容易获得的、能够用来解决问题或做出决策的信息。对可获得性造成强烈影响的一个因素，就是信息的鲜活性。

问题在于，再没有比发自内心的个人见证更鲜活、更引人注目的了，这都是一些已经发生的事或者是真实的事。个人见证的鲜活性常常令其他一些更可靠的信息黯然失色。购物时，我们在不同的品牌前权衡了半天，最后却由于某个朋友或某则广告对于另一产品的推荐，而在最后一刻放弃了自己的选择。买车就是一个典型的例子。在翻看了《消费者报告》中的数千份消费者调查之后，我们终于决定要购买一辆A品牌的车；又参考了几本汽车杂志之后，看到里面的专家们也都推荐A牌子的车，这更坚定了我们的选择——直到在一次聚会上，我们遇到一位朋友，他说他一个朋友的朋友买了一辆A牌子的车，结果是辆残次品，光维修就花了几百美元，而且这哥们决定再也不会买这个牌子的车了。显而易见，这样一个个别案例本不该在很大程度上影响到我们的决定，因为我们是在收集了针对数千名用户所做的调查报告和众位专家的评判之后才决定要买A牌子的车的。然而，我们中究竟有多少人能做到不把这个个别案例看得很重呢？

购买汽车的例子说明，鲜活的个人见证所造成的问题并非心理学领域所独有。鲜活性影响人们决策，这样的例子无论在哪个领域都不难找到。作家迈克尔·刘易斯（Michael Lewis，1997）描述了政治评论家乔治·威尔（George Will）——一个声名狼藉的反对政府干预的人——是

如何在目睹了一场发生在其家门口、导致有人死亡的车祸之后，发表专栏文章呼吁强制使用安全气囊的。

设想一下，一个周五的早上，你在报纸上看到下面这样一个标题："喷气式客机坠毁，413 人死亡"。天啊，你也许会想，多可怕的事故啊！发生了多么糟糕的事情啊！继续设想，在接下来一周的周四，你起床看到报纸写道："另外一场空难，442 人死亡。""哦，不！"你也许会想。"不要再有任何灾难了，多么可怕啊，我们的空运系统怎么了？"然后想象一下——请尽可能地想象——接下来的周五你起床时看到的是："第三起空难悲剧：431 人死亡。"不但是你，整个国家都会抓狂的。联邦政府会被要求尽快调查此事，所有航班禁飞，各种调查委员会如雨后春笋般成立起来，还有海量的法律诉讼被提起。《新闻周刊》和《时代》杂志将会对此作封面报道，它还会占据近期的电视新闻节目的头条。电视纪录片将会对此主题做深度挖掘。躁动和喧嚣是巨大而深远的。

这并不是一个虚构出来的问题，它是真实的。每周都有喷气式客机坠毁。也许不是一架巨型喷气式客机，而是很多小型飞机；或者也不是小飞机，而是小型交通工具，这种小型交通工具叫做汽车。在美国每周都会有超过 350 人死于汽车交通事故（每年超过 19 000 人）(National Highway Traffic Safety Administration，2004），人数足够坐满一架巨大的喷气式客机的了。

每周在高速公路上死于车祸的人数，相当于一架喷气客机的载员数，但我们对此漠然置之。这是因为，"能坐满一架喷气式客机的人死了"这一信息没有通过媒体以一种鲜活的形式传达给我们。因此，每周死于汽车交通事故的 350 人（加上每周死于卡车或摩托车的 330 人），对我们来说不具有鲜活性。我们在餐桌前不会像谈论一架喷气式飞机坠毁并且死了很多人那样谈论这些死于车祸的人。我们不会就汽车出行的安全性和必要性进行争论，但是，如果大型喷气式客机每周都发生坠毁，并且每次都导致 350 人死亡的话，我们就会讨论空运交通的安全性。车祸中死亡的这 350 人不会上新闻，因为他们分布在全国各地，因此对于我们中的大多数人来说只是统计学上的抽象概念。媒体不会为我们生

动地呈现这350名死者，因为他们并不是死在同一个地方。相反，媒体呈现给我们的（有时候）是一个数据（例如，每周350人）。这已经足够引起我们的思考了，但是我们对此毫无反应。与我们生活中的其他任何行为相比，驾驶汽车都是一种极端危险的行为（Galovski，Malta，& Blanchard，2006；National Safety Council，2001；Ross，1999；Sunstein，2002）。然而，关于它的风险和相对应的收益，从未有过全国性的大讨论。这对于住在郊外、需要驾车往返的人来说，是不是一个可以接受的代价？我们从不去问这样的问题，因为问题还没被意识到，而没被意识到的原因就是：代价和风险没有像空难那样以鲜活的方式呈现给我们。

想想下面这个例子的荒谬之处吧。一个朋友开车20公里载你去机场，因为你要乘飞机作一次750公里的旅行。分别的时候，你的朋友很可能会说，“一路平安”，这个临别赠言其实是具有悲伤的讽刺意味的，因为你的朋友在回家的20公里路上死于车祸的风险，要比你飞行750公里的风险高出3倍（National Safety Council，1990）。这就是鲜活性问题，它解释了A对B的安全祝福存在着明显的不合理性，因为恰恰是A正处在更大的风险之中（Sivak & Flannagan，2003）。

科尔（Cole，1998）报道了这样一个民意测验，在环球航空公司（TWA）空难事故后，如果飞机能够提高安全系数的话，很多人都愿意为自己的往返机票多付50美元。同样是这些人，却拒绝为机动车中的安全功能支付50美元，即使那样会提供更加安全的保障。同样地，很少的人能够意识到，当他们驾车10公里去买一张彩票的时候，他们在车祸中丧命的几率要比他们获得头奖的几率高16倍（Orkin，2000）！

这些例子并非只是假设，在“9·11”恐怖袭击事件之后，乘飞机出行的人数锐减，因为人们害怕飞行。当然，人们还要继续外出旅游，而不只是待在家里。他们只是改为其他方式出游——大多数情况下都是自驾车。但是，自驾游比飞行要危险得多，从统计学上讲，注定有更多人因转成自驾游而死亡。事实上，研究者估计，在2001年的最后一个月，有超过300人由于乘坐汽车而非飞机旅行导致死亡（Gigerenzer，

2004）。有一个研究团队能够以一种鲜活的统计来传达出驾驶机动车有多么危险。西瓦克和福兰纳根（Sivak & Flanagan，2003）计算出，如果驾车和乘坐飞机的危险系数是一样的话，那么“9・11”这个级别的事故将会每个月都发生一次！

在媒体的帮衬下，鲜活性误导个人判断的情况在其他领域里也同样广泛存在。一项研究（MacDonald，1990；Cole，1998；Radford，2005）调查了父母最担心他们的孩子遭遇哪种风险。结果显示，父母最担心的是孩子遭绑架，而这一事件发生的概率是 1/700 000。相形之下，父母则不太担心孩子在车祸中身亡的危险，然而这种可能性比遭绑架要高出 100 倍。显然，对绑架的担心大部分是媒体渲染的结果。这项研究结果说明，由于鲜活性效应对知觉的扭曲，“美国父母所担心的竟是一些不大可能发生的事”（MacDonald，1990）。一个研究人员哀叹道：将担心聚焦在“当下流行”的事件上，势必会误导父母的注意力，导致他们“忽视了他们本来能够施加更多影响的方面，例如学业成绩、看电视的坏习惯、吸毒和驾车安全等”（MacDonald，1990）。

作家彼得・鲍耶尔（Boyer，1999）描述了支持持枪自由的议员们如何以同样的方式培养美国人对危险的错误知觉，他们试图将公众的注意力放在“入侵者”破门而入的鲜活案例上。这里的潜台词是：拥有一杆自我保护的枪将会降低你的风险。鲍耶尔（1999）指出了具有讽刺意味的事实，当真实的统计展示了真正的问题在于“枪在好人手里”的时候，枪械制造工业却力图关注“枪在坏人手里”。在这个国家，罪犯并不是导致大多数枪击死亡案的原因。饮弹自尽的人数事实上要远多于被枪杀的人数。大多数死于枪支的人都是枪支走火和自杀——这就是为什么有研究指出，把枪支带回家实际上反而增加了家庭的风险（Miller，Azrael，& Hemenway，2002；Samuels，2004）。

最后，我们对可能患上糖尿病的担心要小于对因感染葡萄球菌而住院的担心，即使前者波及 450 万美国人，而后者每年只有 1500 人而已（Fountain，2006）。这是因为，就个人而言，我们能够对前者做出一些对策（改变饮食或者锻炼），而对后者却无能为力。

单一个案的压倒性影响

人们有这样一种倾向：即使能够获得更为精确的信息，人们的判断也常常受到某个突出例子的影响，对此心理学家已经进行了广泛的研究。威尔逊和布里克（Wilson & Brekke，1994）证明了鲜活性问题是多么具有欺骗性，以及它是怎样影响实际的消费者行为的。他们调查了两类信息如何影响人们选择两个不同品牌的避孕套（品牌A和品牌B）。一类信息是《消费者报告》杂志中的一篇调查报告，另一类则是两个大学生对于偏好的避孕套品牌的观点。首先，威尔逊和布里克调查了一组被试，询问他们更乐于受哪种信息的影响。该组中超过85%的被试都认为比起两个学生的观点，他们更乐于受《消费者报告》文章的影响。研究者随即招募了一组相似的被试，在研究中，被试被告知他们将会免费得到一些他们自己选择的避孕套。被试被告知可以参考以下两类信息中的任何一类或两类：一类信息是在《消费者报告》杂志中的一篇调查报告，另一类是两个大学生的观点。尽管此组中只有不到15%的人愿意接受两个大学生的观点，但还是有77%的人同时询问了两类信息。很显然，被试无法抵御他人见证的诱惑，尽管他们不认为自己会受其影响，但他们事实上却被影响了。当被试同时选择了解两类信息并且这两类信息相互冲突时，接受学生推荐意见的人数比接受《消费者报告》推荐意见的人数多31%。

另外一个关于人们如何对鲜活的轶事信息做出不同反应的例子，来自于在20世纪60年代中后期媒体对越战的报道。随着战事的拖延，美军的死亡人数仿佛无休止地增加，媒体开始报道当周美军死亡的人数。一周接一周地过去了，这个数字在200至300之间徘徊，公众似乎已对这种报道习以为常了。然而，某杂志用几个版面的篇幅连续刊登了前一周阵亡者的个人照片。这时公众非常具体地看到了在这样一个有代表性的一周内逝去的大约250个鲜活的生命。结果，此举导致了大规模的、针对这场代价巨大的战争的抗议声浪。250张照片所产生的影响是每周数字报道所远不能及的。但是作为一个社会成员，我们应该克服这种不

相信数字、必须亲眼目睹才去相信的倾向。绝大多数影响我们社会的复杂因素都只有靠数字才能捕捉。只有当公众学会像重视图像材料一样重视以数字形式表达的抽象材料时，公众自己的立场才不会像屏幕上闪过的最新图像那样变化无常。

2004 年，一档叫做《晚间在线》的电视节目在伊拉克战争一周年之际，公布了在这场战争中死亡的 700 多名战士的名字和照片，在这一时刻，历史又重演了。这一做法与该档节目在“9 · 11”事件一周年之际播放受害者的姓名和照片的套路完全相同，当时这些照片的播放都征得了受害者家属同意。然而，死亡士兵的照片还是引发了战争支持者的抗议。有些人控诉节目主持人泰德 · 考佩尔对这场战事抱有敌意，但是这些指控显然瞄错了对象，因为考佩尔并不反对这场战争。相反，考佩尔说，“你们中的一些人深信我是反对战争的，其实我并不是，但这不是重点。我只是反对那种一直以来的幻觉，即认为战争仅牺牲少数人，不会连累我们余下的人”（CNN.com，2004）。战死的人数并非没有被报道，这 700 多人死亡的消息日复一日地出现在这个国家的每一份报纸上。但是争论的双方都知道，公众尚未对那些数字进行“加工”——没有计算代价，是因为那些数字还太过抽象。双方都知道很多人在看过这些照片之后，都会从头对这些信息进行加工，并开始真正在意战争的代价。

不仅公众受到鲜活性问题的困扰，在心理学和医学领域，有经验的临床从业者一直都在努力摆脱个别案例的压倒性影响给他们的决策带来的阴影。作家弗兰辛 · 卢索（Russo，1999）描述了弗吉尼亚大学的肿瘤专家威利 · 安德森面对的两难境地，安德森一直提倡控制实验，并会定期招募一些病人来做有控制的临床测试。但是他仍旧纠缠于自己对突出个案的反应，那些鲜活的个案对他的决策产生了影响。尽管他相信科学，但仍承认“当真实的人眼巴巴地看着你的时候，你将被他们的期望以及自己对他们期望的期望所包围，这确实非常困难”（p. 36）。但是安德森知道，有时对他的病人来说，最好的办法就是忽略“看着你的那个真实的人”，并且遵循最佳证据的指示。最佳证据来自于有控制的临床

试验（将在第 6 章表述），而不是看着你的那个人的情感反应。

评估证据的时候，鲜活性问题是一个我们都要面临的难题。并且，在这样一个充斥着媒体影像的环境中，对于社会来说，不受影像支配而基于有效的证据来解决自身的问题变得越来越困难了。作家巴瑞·格拉斯纳（Glassner，1999）讲述了一个特别相似的例子。在 1995 年的一场奥普拉秀中，围绕着某一次外科手术而展开讨论（这里不具体说出外科手术的名字，以免渲染鲜活性效应）。这种手术引起了一些争议，因为一些人声称在手术过程中受到过伤害。来自梅奥临床医院、哈佛大学和密歇根大学的研究证据都显示，这种手术总体上是没有危险性的。就在此时，一位声称受到过伤害的妇女从观众席里跳出来并喊道：“我们就是证明，我们这些坐在这儿的人就是结论！”（Glassner，1999，p. 164）。你认为哪种说法会让数百万电视观众记忆犹新——是来自梅奥临床医院的研究，还是大喊自己就曾受到过伤害的那位妇女？

总之，过于依赖见证证据的问题一直存在。此类证据的鲜活性常常掩盖了更加可靠的信息，并且混淆视听。心理学教师担心的是，仅仅指出依赖见证证据的逻辑谬误，并不足以让人们从一个更深的层次理解这类数据的缺陷。我们还能做些什么呢？还有什么其他的方法能让人们理解这个概念吗？幸运的是，我们还有一个法宝——一种与学术方法略有不同的方法。这种方法的本质是以鲜活性来对付鲜活性，是以一种“以彼之道，还施彼身”的方法对付见证证据，让见证用自身的荒谬来击溃自己。这个方法的实践者，就是独一无二、毋庸置疑的“了不起的兰迪”！

了不起的兰迪：以彼之道，还施彼身

詹姆斯·兰迪（James Randi）是一位魔术师，并且是个多面手，他曾经被麦克阿瑟基金会授予过“天才”奖。多年来，他一直尝试着教公众学会一些基本的批判性思维的技巧。“了不起的兰迪”（Amazing Randi，他的艺名）通过揭穿“通灵”骗术和庸医疗法来达到教育公众

的目的。尽管他拆穿了很多魔术和伪装的所谓“通灵术”,但最为著名的，还是他拆穿 20 世纪 70 年代通灵术超级明星尤里·盖勒（Uri Geller）的把戏的那一回。盖勒靠吹嘘通灵术红透荧屏，他对媒体的蛊惑程度简直可以用无以复加来形容。各大洲的报纸、电视节目和主要的新闻杂志对他争相报道（盖勒仍健在，还在写书；Radford，2006）。兰迪发现并揭露了盖勒经常表演的通灵术“绝活”其实不过是些普通和简单得令人难以置信的魔术把戏，包括使勺子和钥匙弯曲、使钟表开始走动等等，这些对于一个优秀的魔术师来说，简直就是家常便饭。自从盖勒被拆穿以后，兰迪继续将他那非凡的才智用于维护公众的知情权，他不断揭露超感官感知、生物节律、超自然力、通灵外科手术、天外来客、漂浮术以及其他伪科学的谬误，帮助公众了解真相（Randi，1983，1995，2005；Sagan，1996）。

兰迪的另外一个兴趣就是去证明，对于任何一个荒谬的事件或无中生有的言论而言，获得见证是多么地容易。他的手法就是，让人们掉进其见证所编织的陷阱里。兰迪常常使用脱口秀这一理想的美国文化载体来实现他的目的。他经常作为嘉宾出现在节目中，但不以真名示人。在几年前纽约的一个节目中，他对观众说，今天早晨开车经过新泽西的时候，他看到一个橙色的 V 形物体飞过头顶飘向北方。几秒钟之内，正如兰迪所说，“电台的接线总机像一棵电子圣诞树般闪烁起来”。一个接一个的目击者打电话过来证实这一奇异的景象。可他们运气不好，这一“景象”只是兰迪想象出来的,但是打电话的人提供了许多兰迪“忽略了”的细节，包括其实有多个“飞碟”飞过。这个小小的把戏证明，个体关于“某事发生”的报告是多么不可靠。

在加拿大的温尼伯市，兰迪在一个广播节目中以“星相学家”的身份出现。节目一周前，听众被要求提交他们的笔迹样本和出生日期。有三个人被甄选出来，并且进行电话连线，这样兰迪就可以“解读出”他们的性格特征。他大获成功，三位听众在 1~10 分的准确性量表上给出的评分分别为 9、10 和 10。兰迪在节目的最后向听众解释了他的秘诀。他其实只是逐字逐句地读出了占星师希德尼·奥马尔在最近的一个电视

节目中给三个观众的“解读”而已。

在另外一个广播节目中，兰迪揭示了另外一种伪科学——生物节律能够如此流行的原因（Hines，1998，2003）。一位听众同意每天都记日记，并将日记与一份特别为她准备的两个月的生物节律表做比较。两个月以后，她打回电话告诉听众：生物节律绝对不是盖的，因为节律表预测实际行为的准确率超过了90%。兰迪不得不把他的秘书所犯的一个愚蠢的错误告诉给这位听众，秘书错误地将本该发送给另外一个人的节律表发给了她，而不是她自己的。然而，这位妇女还是同意看一下真正属于自己的表格是怎样的，于是，又一份表格立即发送给了这位妇女，并且请她再打电话过来。几天后，这位妇女带着解脱感打进电话，说她自己的表格也同样十分准确——事实上更为准确。在下一期节目中，大家发现，另一个错误发生了。这位妇女收到的是兰迪秘书的节律表，而不是她自己的！

兰迪的生物节律和占星术小把戏，其实是一种被命名为巴纳姆效应（Barnum，著名的嘉年华和马戏团的团主，提出了“每分钟都会有人上当受骗”的说法）现象的范例。这一效应曾被心理学家广泛地研究（例如，Dickson & Kelly，1985），研究者发现，大多数成年人都会认为泛化的个性总结都是准确的，并且都是对自己独特的描述。这里有一个来自谢尔默（Shermer，2005，p. 6）的例子：

> 你是一个非常体贴的人，总是及时地帮助别人。但是也有一些时候，你会发现你有一点点自私……有时候你太忠于自己的感受以至于会暴露过多的自己。你善于思考，并且对任何事情，在改变想法之前都希望看到证据。如果你处在一个陌生的环境下，你会非常小心，直到你看清楚发生了什么事情，然后才会充满信心地行动……你知道怎样做一个好朋友，你懂得训练自己，所以在别人看来你都在掌控之中，但其实有些时候你是缺少安全感的。你希望在人际关系中比现在更受欢迎，更加自如。你面对世界表现得很有智慧，这种智慧来源于艰难的体验而非书本学习。

大多数人都发现，这个总结是对其个性非常准确的概括，但是很少有人自发地意识到大多数其他人也同样认为它描述了他们自己！许多众所周知的语句和措辞（如这个例子）使很多人认为适用于他们自己。谁都能够将其作为一个个人化的心理“分析”提供给“顾客”，而这些顾客常常会为个人化的“性格解读”的“准确性”而感到震惊，却不知道其实每个人的解读都是一样的。当然，巴纳姆效应正是手相学和占星术的基础（Kelly，1997，1998）。巴纳姆效应还可以证明产生见证有多么容易，以及为何见证毫无价值。

这就是詹姆斯·兰迪运用这些小把戏努力想要达到的目的——给人们好好上一课，告诉人们见证证据是没有价值的。他不断地证实，形成有利于虚假主张的见证是多么容易。正是这个原因，用见证来支持自己提出来的特定理论是毫无意义的。检验一个主张时，只有来自于有控制的观察中的证据（第 6 章中将会描述）才是足够充分的。

见证为伪科学打开方便之门

有时候有人会说，类似刚才所讨论的种种伪科学，只不过是人们给自己找乐子的一种方式，无伤大雅。再者说，我们又何必较真呢？不就是有几个人在异想天开，而另外几个人从中赚点儿小钱吗？

然而，对此问题进行一番彻底的考察就不难发现，伪科学的盛行对社会的危害比人们想象的要大得多。在一个复杂的、科技化的社会中，一些能够影响千万人的决策会为伪科学的影响推波助澜。也就是说，即使你并不认同这些伪科学的观念，你也可能受到这些观念的影响。例如，大银行和一些 500 强企业雇佣笔迹学家来做人事选拔（Sutherland，1992），即便大量的证据表明，笔迹学在实现这一目的方面是没有作用的（Ben-Shakhar，Bar-Hillel，Blui，Ben-Abba，& Flug，1989；Neter & Ben-Shakhar，1989）。伪科学的笔迹学指标在一定程度上使雇主忽视了其他更有效的选拔标准，导致的结果是经济上的零效益和对一些人的

不公平待遇。如果仅仅因为笔迹中有一个连写的“小圈圈”，就让你丧失了获得一份你很心仪的工作机会，你会作何感受？或者，你求职被拒，只是因为一个“通灵师”从你的“气场”中看到了一丝扰动，你又会作何感受？事实上，这类事情的确发生在一些人身上，一些公司会花钱请人为求职者进行“通灵分析”。例如，苏姗·金（Susan King）是一个所谓的“通灵大师”,公司花钱请她运用“读心术”来为人事决策提供帮助。“尽管一些雇主会请她来参与最后一轮面试，或者在随后的鸡尾酒会上让她去观察一些入围的候选人”，但她宣称她甚至不需要见到这些申请者——她可以从照片或是他们的姓氏中发现问题（Kershaw，1991）。在这样一个竞争激烈的经济时代，这是你所期望的决定自己能否获得一份工作的方式吗？

不幸的是，这样的例子绝非凤毛麟角（Shermer，2005；Stanovich，2004；Sternberg，2002）。当伪科学的观念渗透于整个社会的时候，我们都以不同的方式受到影响——即使我们并不认同这些信念。例如，警局雇通灵师协助办案（Marshall，1980），即便研究表明这一举动是没有任何效果的（Hines，2003；Rowe，1993）。ABC电视网络的制作人居然花钱请好莱坞的通灵师帮助他们确定节目内容，要知道电视网络可是当今社会最具影响力的通信科技（Auletta，1992，p.114）。最令人大跌眼镜的是，有个占星师曾受雇于里根时代的白宫，专门为“总统演说的时机、穿着打扮、与州长会面、飞机的飞行时刻表甚至讨论的议题等”这类事项提供建议（Johnson，1991，p. 454）。

如今，类似占星术这样的伪科学是一项巨大的产业，涉及报纸专栏、广播节目、图书出版、网络、杂志文章以及其他各种传播渠道。星相学杂志的发行量要比很多正规的科学杂志大得多。据美国众议院老龄化问题委员会估算，浪费在医疗骗术上的钱已经达到数十亿美金。简而言之，伪科学是个油水颇丰的行当，数以千计的人靠公众的盲信盲从而获得收益。

医学领域中的伪科学话题具有借鉴意义，因为在抨击伪科学，以及把正规的与不正规的医疗实践划清界限方面，医学界的各类组织都比心

理学界表现得更为激进和勇猛。下面就让我们看看由关节炎基金会出版、曾被美国众议院老龄化问题委员会所引述的一套识别不道德药品推销员的指南：

1. 他或许会提供一种用于治疗关节炎的“特别的”或“秘密的”处方或设备。
2. 他会做广告，用的都是“个案史”和“满意患者”的见证。
3. 他或许会承诺（或者暗示）能够快速或轻松见效。
4. 他也许会声称知道关节炎的成因，并且说能够“清除”你体内的“毒素”，同时促进你的健康。他或许会说外科手术、X 光和医师所开的处方是没有必要的。
5. 他或许会指责“医学体制”故意阻碍了进步，或者迫害了他……但是他不允许他的方法以已有的或已获证明的方法来验证。（U. S. congress，1984，p. 12）

这份清单同样可以作为识别带有欺骗性的心理学疗法和理论的指南。在这里，请注意第 2 条，这正是本章关注的焦点。同时注意，第 1 条和第 5 条论证了之前所讨论过的一个观点：科学是公开的。除了宣扬见证叙述作为“证据”，伪科学的从业人员经常以指责他人有意要压制他们所获取的“知识”，以此试图逃避“公开可证实”这一科学的标准。这样，他们就有借口带着他们的“研究成果”直接走进媒体，而不是通过正规的科学出版程序将这些成果公诸于世。通常，这种伎俩在心理学领域中更为成功，因为相比于其他科学领域，媒体通常对心理学的正常科学机制缺乏尊重。记住这一点，这很重要（在第 12 章中我们将会深入讨论这个问题）。媒体从来不会考虑去报道物理科学中未经证实的主张，但如果这类主张是心理学方面的，就会被当做正规的心理学话题加以报道，因为新闻记者早就被宣扬“心理学里没有规则”的伪科学洗了脑。然而，消费者必须意识到，电视和纸质媒体只要认为读者有需要，就会报道心理学领域中任何出格的主张，无论这些主张与已有证据之间是多么矛盾。最终的受害者是公众。

有关神奇疗法的宣传助长了人们的错误希望，当希望破灭时，会给人们造成心理上的致命打击。在我关于这个问题所掌握的例子中，其中最卑劣的案例之一就是一篇来自杂货店小报的文章，标题冠以“通灵师展示盲人如何通过超感官知觉看见东西”。人们可能会由于身陷伪科学的重重包围而无法获得真正科学的知识，通灵外科手术的倡导者暗中怂恿人们把钱花在骗人的疗法上，并且忽视对患者有帮助的传统的“非通灵的”的医疗手段（Angell & Kassirer，1998）。在一个关于医疗保健欺诈行为的市民大会上，“密歇根反健康欺诈顾问委员会”的主席展示了一则骗人的、治疗癌症的小册子，上面写着“本产品不能与其他癌症疗法同时使用”（“听上去像真的”，1990）。类似案例已造成的损害是无法估算的。

心理学家越来越关注医学骗局在互联网上的蔓延，以及它对健康带来的损害。麦克斯·考皮斯（Max Coppes）博士不得不给《新英格兰医学杂志》写了一封信，警告人们注意医学中的伪科学所带来的危害（Scott，1999）。他描述了一个 9 岁女孩的案例，这个孩子在经历癌症手术之后，如果接受化疗的话，将会有 50% 的机会可以多活 3 年。但她的父母找到一种未经验证的、利用鲨鱼软骨的偏方来代替化疗。小女孩在 4 个月后失去了生命。

当我正在讲述这个话题的时候，经常有人会针对我的演讲提出非常中肯的问题：“你不也是正在用生动的个案来阐述你的观点吗——这种做法难道不正是你所反对的吗？”这个问题问得好，并且它让我有机会详细阐述本章中包含的一些论点间的微妙之处。这个问题的答案是肯定的，我运用了生动例子来阐述观点。但是，是为了阐述观点，而不是为了证明观点。这里的关键是要区分两点：主张的提出和主张的交流。对于每个主张，我们都能问这样一个问题：它是不是基于鲜活的见证？这会产生四种可能的情况：

1. 一项主张基于鲜活的见证，同时依靠鲜活的见证来交流
2. 一项主张基于鲜活的见证，同时不依靠鲜活的见证来交流

3. 一项主张基于证据而非鲜活的见证，同时依靠鲜活的见证来交流
4. 一项主张基于证据而非鲜活的见证，同时不依靠鲜活的见证来交流

本章中的一些讨论属于第 3 种情况：一项主张基于证据而非鲜活的见证，同时依靠鲜活的见证来交流。例如，我引用了很多非见证的证据贯穿整章，就是为了说明：个案研究的证据不能用于建立因果性结论，鲜活的例子在人们的判断中被赋予了过高的权重，伪科学的代价巨大等等。对于这些主张中的每一项，我都标出了引证和参考文献。尽管如此，出于交流的目的，我使用了一些鲜活的案例，将注意力吸引到这些主张上，并让它们给人们留下深刻的印象。关键的一点是，支持这些主张本身的并不仅仅是鲜活的见证。比如，我曾使用一些鲜活的例子来阐述"鲜活的例子在人们的判断中被赋予了过高的权重"这一事实，但是这一主张的证据包含在我所引用的经过了同行评议的科学证据之中（例如，Kunda，1999；Lassiter et al.，2002；Nisbett & Ross，1980；Sinaceur，Heath，& Cole，2005）。

回到这部分的主要观点上并做个总结吧：伪科学的传播所造成的代价是巨大的。需要搞清楚哪种类型的证据能够揭示某种现象中蕴含的道理或理论是否可信，如果搞不清楚这一点，就会大大有利于伪科学的传播。由于见证叙述可以为任何主张提供唾手可得的支持，以及自身所具备的冲击力，见证打开了通往伪科学的大门。对于心理学信息的消费者来说，对它们保持警惕应当是头等大事。在接下来的几章中我们将会看到，在证实某种主张的合理性时，究竟需要哪些类型的证据。

小　结

个案研究和见证叙述在心理学（以及其他科学）研究的早期阶段是有用的，因为此时，寻找有趣的现象和待研究的关键变量很重要。虽然个案研究在早期的、理论形成前的阶段是有用的，但在研究的后期，当

对理论进行检验之时，个案研究就毫无用处了。这是因为，作为一个孤立现象，个案研究的结果遗漏了太多其他可能的解释。为何个案研究和见证证据对于理论检验来说是没有用的？要想理解这一点，就需要想一想安慰剂效应。安慰剂效应是指，无论疗法是否包含了有效的成分，人们都倾向于报告任何疗法都对他们有效。安慰剂效应的存在，催生了许多关于疗效的见证叙述，致使对一种心理（或医学）疗法效果的证明成为“不可能的任务”。原因就在于，无论治疗手段是什么，安慰剂效应都会使人们提出证实其疗效的个人见证。

尽管见证证据在检验理论的时候是无用的，但心理学研究指出，由于鲜活性效应，这类证据经常被人们过分地倚重：对于更为生动、并因此在记忆中更易提取的证据，人们会赋予其过高的权重。对大多数人来说，见证证据就是一种格外生动和鲜活的信息，因此，人们在验证某一心理学主张的合理性时，会过度依赖这类证据。事实上，理论主张是否合理，是不能用见证叙述和个案研究的证据来判定的。

5

相关和因果

——用“烤箱法”避孕

几年前，在中国台湾地区曾开展过一次大规模的研究，目的是调查哪些因素和人们对避孕工具的使用有关。一个由社会学家和内科医生组成的大型研究团队收集了有关环境和行为变量方面的大量数据。研究者比较感兴趣的是，哪种变量能够最准确地预测避孕方法。数据收集上来之后，研究者发现，有一个变量和使用避孕工具的相关最强，这就是：家庭中家用电器（烤箱、风扇等等）的数量（Li，1975）。

这个结果恐怕不会促使你提出这样的建议：在高中发放免费的烤箱以解决青少年的怀孕问题。但是，你为何不会有这样的想法呢？电器和避孕工具使用之间的相关性很高，在众多被测量的变量中，这个变量是唯一最准确的预测因子。我希望你的回答会是：问题关键在于这两个变量间关系的“性质”而非“强度”。开展“免费烤箱计划”预示着这样一种观念：烤箱导致人们使用避孕工具。而实际上我们会将这种建议视为一种荒唐的方案，至少在上面所举的这个显而易见的例子中，我们会认识到，这两个变量可能有相关，但不是因果关系。

在这个例子中，我们可以猜想，这种关系之所以存在，是因为“避孕工具的使用”和“家庭中家用电器的数量”这两个变量通过与这两种

变量都相关的其他变量联系起来。教育可能会是中介变量之一。我们知道，教育水平与避孕工具使用和社会经济地位都有关系。现在我们所需要的就是这样一个事实：经济水平高的家庭会拥有更多的家用电器，我们都会有这样的联想。当然，其他的变量也可能会在二者的关系中起到中介作用。但是，无论“家用电器的数量”和“避孕工具使用”之间的相关有多么强，这种关系都不能说明它们之间存在因果关系。

避孕方法的例子很容易让我们理解这一章的主旨：有相关，并不意味着必然有因果关系。本章中我们将会讨论阻止我们做出因果推论的两大问题：第三变量问题和方向性问题。我们还将会讨论选择性偏见是如何导致第三变量问题的。

相关性证据的局限性并不都像“烤箱”例子那样容易被识别。当因果关系对我们来说显而易见时，当我们抱有根深蒂固的偏见时，或者当我们的解释被理论定势所主宰时，就会很容易地把相关当作因果的证据。

第三变量问题：戈德伯格与糙皮病

在 20 世纪初期，数以万计的美国南部民众罹患并死于一种叫做糙皮病的疾病（大约每年 100 000 人死亡）。糙皮病被认为是由一种不明微生物引发的传染性疾病，其主要症状是头晕、嗜睡、溃疡、呕吐和严重腹泻（Chase，1977，p. 205）。此后，许多来自全国糙皮病研究学会的医生都认同这样的证据：糙皮病和卫生条件有关。这并不令人吃惊。家在南卡罗来纳州斯帕坦堡的人们似乎总是远离糙皮病的困扰，因为他们有自来水管道和良好的污水处理设施。这种相关恰好验证了这样的观点：由于糟糕的卫生条件，传染性疾病是通过糙皮病患者的排泄物传播开来的。

一位叫约瑟夫·戈德伯格（Joseph Goldberger）的医生对这种解释非常怀疑，在美国公共卫生部部长的指示下，戈德伯格针对糙皮病开展了许多研究。他认为糙皮病是由于营养不均衡的饮食引起的，简而言之，

是美国南部普遍的贫困造成的。许多的患者赖以生存的都是高碳水化合物、蛋白质含量极低的饮食，如很少量的肉类、蛋类、牛奶，以及大量的谷类、燕麦和玉米粥。戈德伯格认为污水处理条件和糙皮病之间的相关在任何一个方面都无法反映因果关系（和烤箱控制生育的例子一样）。他认为根本原因在于，拥有清洁管道的家庭通常也都是经济状况良好的家庭，经济上的差异也会反映在他们的饮食上，经济状况好的家庭在其饮食中包含更多的动物蛋白。

但是，请等一下！为什么戈德伯格的因果推断就一定是对的呢？毕竟，两派人马都是坐在那里，根据相关数据推论什么才是造成糙皮病的原因的。为什么医学会的医生们不能说戈德伯格的相关同样也是误导性的呢？为什么戈德伯格能够推翻别人的假设———种微生物通过糙皮病患者的排泄物传播，而这种传播是因为不完善的污水处理设施造成的？戈德伯格对糙皮病的判断还涉及一个小细节，这个细节我刚才没说：戈德伯格吃下了糙皮病患者的排泄物。

为什么戈德伯格的证据更好

戈德伯格有一类这样得来的证据：研究者不仅观察相关性，还靠真正地操纵关键变量来收集数据（有关控制操纵，将在下一章进一步讨论）。这种方法经常要创造一些通常极少会自然出现的条件——说戈德伯格设计的特殊条件不会自然出现，无论怎样强调都不会过分。

戈德伯格确信糙皮病是不会传染的，也不会通过患者的体液传播，他给自己注射了一名患者的血液，还吃进一名患者喉咙和鼻子内的分泌物。此外，

> 他还选择了两个病人：一个有皮癣症状，另一个有腹泻。他从皮癣处刮掉鳞屑，然后和该病人的 4 毫升尿液混合到一起，然后再加上相同数量的液体排泄物，最后与 4 小撮面粉揉在一起做成小药丸。戈德伯格、戈德伯格的助手以及戈德伯格的妻子自愿服下这些

药丸。(Bronfenbrenner & Mahoney, 1975, p. 11))

无论是戈德伯格，还是其他的志愿者，都没有染上糙皮病。简言之，戈德伯格创造了这个传染疾病可能传播的所有条件，结果平安无事。

戈德伯格对其他人提出的因果机制进行了操作，结果显示该机制是无效的，尽管如此，对他自己提出的因果机制进行检验仍然非常必要。戈德伯格选择了来自密西西比州监狱农场的两组犯人，这些人都是没有患糙皮病的，并且都是自愿参加实验。其中的一组人被给予高碳水化合物、低蛋白质的食物，这种类型的食物是戈德伯格怀疑引起糙皮病的原因。另一组被试被给予(营养成分)更均衡的饮食。5 个月后，低蛋白质的这一组患上了糙皮病，而另一组却没有丝毫的患病迹象。戈德伯格的理论遭到了一些人的反对，这些人出于政治动机而否认贫困的存在。经过长期的抗争，戈德伯格的假设终于被人们所接受，因为他的假设与实验证据的契合程度是其他任何假设所不能比拟的。

糙皮病的历史说明，如果依据相关研究来制定社会和经济政策，必将使人类付出惨痛的代价。但这并不是意味着我们永远不要使用相关研究的证据。恰恰相反，在许多场合，我们必须用到相关(见第 8 章)，而在某些情况下，只要有相关就够了(例如，当我们的目标是预测而不是决定原因的时候)。科学家们经常不得不使用不充分的知识来解决问题。重要的是，我们在运用相关性证据的时候要谨慎小心。像“糙皮病 – 污水”这样的案例，在心理学研究的每个领域内都频频发生。这个例子也揭示了“第三变量问题”：事实上两个变量之间的相关——这个例子中是糙皮病的发病率和污水处理条件——并不意味着这两个变量之间有直接的因果关系，相关之所以产生，是因为这两个变量都分别与第三变量相关——这里是饮食——而这个变量没有被测量。像这种污水处理条件和糙皮病之间的相关，我们通常称之为“虚假相关”：相关的产生不是因为两个变量之间存在一个可以测量的直接的因果联系，而是因为这两个变量都与第三变量相关。

下面我们来看一个发生在现实生活中的例子。多年以来，有关公立

学校和私立学校教学质量的争论甚嚣尘上。从这场争论中得出的一些结论，很生动地展示了从相关证据推出因果关系的弊端。私立学校和公立学校的好坏是一个实证性问题，可以使用社会科学中的调查研究方法来辨别真伪。但是，这并不意味着只要这个问题是个科学问题、有可能获得解决，就是一个非常简单的问题。所有鼓吹私立学校优越性的人都潜在地意识到这一点，因为他们在维护自己的观点时，常常引用这样一个经验性的事实：私立学校学生的成绩要好过公立学校。尽管这个事实无可辩驳——各种研究中有大量一致的教育统计数据，但问题在于，用这些学生的成绩数据就推出结论，即私立学校的教育本身导致了较高的分数，这么做是否合适？

考试成绩是许多不同变量的函数，这些变量彼此之间又是相关的。为了评估公立学校和私立学校的好坏，我们需要进行更为复杂的统计，而不仅仅是学校类型和学业成就之间的相关。例如，学业成就和家庭背景中许多不同指标都有关系，如父母的教育程度、父母的职业、社会经济地位、家中藏书的数量以及其他一些因素。这些特征都与是否把孩子送到私立学校有关系。因此家庭背景是一个潜在的第三变量，可能会影响到学业成就和学校类型之间的关系。简言之，学业成就可能和学校质量没有任何关系，而结果可能是：家境优越的孩子学习更好，更有可能进入私立学校。

幸运的是，还有许多复杂的相关统计方法，例如*多元回归*、*偏相关*、*路径分析*（统计学的发展部分要归功于心理学家），这些复杂的统计方法能够去除其他变量的影响、提出公因子或定义协变量之后重新计算两个变量之间的相关。来自杜克大学的艾利斯·佩奇和蒂莫西·凯斯（Ellis Page & Timothy Keith，1981）则使用更为复杂的统计技术，分析了一系列关于高中生教育的统计数据，这次统计数据的收集是在国家教育统计中心（NCES）的资助下进行的。他们发现，当反映学生家庭背景和一般智力能力的变量被排除后，学业成就和学校类型之间几乎就没有一点关系了。其他研究者也确认了他们的研究结果（Berliner & Biddle，1995；Carnoy，Jacobsen，Mishel，& Rothstein，2005）。

因此，很明显，鼓吹私立学校能够提高教育成就，就跟讨论节制生育需要用“烤箱”一样没什么分别。学业成就和私立学校相关，不是因为任何直接的因果机制，而是因为私立学校中学生的家庭背景和一般认知水平与那些进入公立学校的学生相比是不一样的。

这些较为复杂的相关统计方法，能够排除第三变量的影响，但并不总是会削弱原有相关的强度。有时候，在排除第三变量之后，两个变量之间的原有相关仍然存在，这个结果本身就能提供一些信息。这样的结果说明，原有相关并不是由第三变量所导致的虚假相关。当然，并不排除其他变量也会导致虚假相关。

托马斯、亚历山大和埃克兰德（Thomas，Alexander，& Eckland，1979）提供了数据分析方面的一个好例子。这些研究者发现，高中生是否进入大学和这个学生的家庭社会经济地位有关。这是一个重要发现，足以动摇我们这个社会的核心价值——实现目标靠的是个人能力。它表明，一个人的成功取决于这个人的经济地位。但是在下这个结论之前，我们必须首先考虑一下其他假设。这就是：升入大学和社会经济地位之间的相关是一种假象。其中一个非常明显的第三变量就是学业能力，它可能与升入大学和社会经济地位都有关系，如果这个变量被排除出去，这两个变量之间的相关就会消失。在学业能力被排除后，研究者计算出的数据发现，升入大学和社会经济水平的相关仍然显著。因此，高收入阶层的孩子更容易进入大学不能完全归因于学业能力的不同。当然，这个发现不能排除这种可能性：其他一些变量导致了升入大学和社会经济水平之间的相关，但是能够用这样一种再分析来排除学业能力对两者相关的影响，这本身就在理论及实践方面具有很重大的意义。

安德森等（Anderson & Anderson，1996）描述了他们是如何来检验关于暴力的地区差异理论的，他们通过检验一系列不同的理论看其是否能够对所收集的数据做出解释。他们采用偏相关技术来进行此项研究。曾有研究表明美国南部地区的暴力犯罪高于北部地区，他们检验了“热假设”——令人不适的高温增强了侵犯性动机和攻击性行为（p. 740）。他们发现城市平均气温和暴力犯罪率之间存在相关，这并不令人奇怪。

但是从统计上控制一些变量，如失业率、个人平均收入、贫困率、教育程度、人口规模及其他一些变量之后，气温和暴力犯罪之间的相关仍然显著。这就使得“热假设”理论的可信度大大提高了。

方向性问题

如果能够采用某种方式操纵变量，并能够因之做出科学的因果推断，就没有理由仅凭相关证据做出因果推论。而让人苦恼的是，当涉及心理学主题时，仅根据相关就得出结论的做法却是普遍存在的现象，在心理学知识对解决社会现实问题愈发重要的今天，这种倾向所造成的损失也与日俱增。在教育心理学界，一个广为人知的例子很好地诠释了这一点。

自从 100 年前关于阅读的科学研究开始以来，研究者们就知道，眼动模式和阅读能力之间存在着相关。阅读能力差的人，其眼动轨迹是不规则的，表现为更多的回扫（从右向左的运动），在每一行上的注视时间（停顿）更长。基于这种相关，一些教育工作者假设，眼球运动技能的缺失是造成阅读问题的原因，因此许多“眼球运动训练计划”在小学生中展开和实施。在查明这一相关是否真的说明“不规则的眼球运动会导致低下的阅读能力”之前，这些训练计划已经开展了很长时间。

现在已经清楚了，眼球运动与阅读能力的相关反映了一种与之前所想象的完全相反的因果关系。不规则的眼动并不导致阅读障碍（Rayner, 1998），相反，是缓慢的单词识别和理解困难导致了不规则的眼动。当教会儿童有效地识别单词和更好地理解文字后，他们的眼动轨迹变得平顺了。训练儿童的眼球运动和提高其阅读能力是没有关系的。

最近十几年以来，研究者们已经明确指出，文字解码和语音加工方面的语言问题是阅读障碍存在的根源（Rayner，Foorman，Perfetti，Pesetsky，& Seidenberg，2001、2002；Snowling & Hulme，2005；Stanovich，2000；Vellutino，Fletcher，Snowling，& Scanlon，2004），而几乎没有眼动模式导致阅读障碍的案例。但是，如果到大部分中等规

模以上的学校的储藏室里仔细翻一翻，都能找到布满灰尘的眼球运动训练仪器，这表明数以千计的买设备的钱被浪费了，这就是把相关视为因果证据的后果。

第 1 章中讨论过一个类似的例子。在教育和社会服务领域里有一个非常流行的观点：学业成就问题、药物滥用、青少年怀孕以及其他一些问题行为都是低自尊造成的。这一说法假定，此因果关系的方向很明显：低自尊导致行为问题，高自尊带来高的学业成就和其他领域的成绩。这种方向性因果关系假设为许多提高自尊的教育计划提供了动力，这个问题和眼球运动那个例子是一样的：仅仅因为存在相关就推出一个方向性的因果假设。事实证明，就算真的存在因果关系，自尊和学业成就之间的关系更可能呈相反的方向：高学业成就（包括生活中其他方面）导致了高自尊（Baumeister，Campbell，Krueger，& Vohs，2003；Stout，2000）。

到目前为止，我们的讨论主要围绕变量间相关所涉及的两种误区。其中一种叫做*方向性问题*，已经通过眼球运动和自尊的例子进行了阐述。当变量 A 和变量 B 之间存在相关时，在断定 A 的变化引起 B 的改变之前，我们必须清楚因果关系的方向可能是相反的，即从 B 到 A。第二种是有关第三变量的问题，此问题已经通过糙皮病的例子（以及烤箱—节育和私立学校—学业成就的例子）加以论述。两个变量之间的相关并不能预示任何方向上的因果，因为当这两个变量都和第三变量相关时，该相关还是会出现。

选择性偏差

在一些情境下，虚假相关很容易出现。这也正是选择性偏差非常容易出现的原因。“*选择性偏差*”这个术语指的是特定主体和环境变量之间的关系，当不同生理、行为、心理特点的人们选择不同类型的环境时，就有可能出现选择性偏差。选择性偏差造成环境特征和行为 - 生物特征之间的虚假相关。

让我们通过一个例子来了解选择偏差是如何产生虚假相关的。请快速说出一个州名，在这个州里，由呼吸系统疾病导致的死亡率高于平均水平。当然，答案之一是亚利桑那州。什么？等等！难道亚利桑那州没有清洁的空气吗？难道洛杉矶的烟雾弥漫得如此之远？难道凤凰城的郊区环境已经变得那么差了吗？不是，肯定不是！让我停下来想一想。可能亚利桑那州的确有清洁的空气，可能患有呼吸疾病的人都愿意搬到这里，然后他们死在了这里。这样就对了。如果我们不够认真，就会出现上面所说的那种情形：我们可能会受到误导，认为是亚利桑那州的空气害死了这些人。（亚利桑那州在美国西部，以地广人稀、气候干燥、空气清洁著称——译者注）

但是，选择性偏差并不总是那么容易辨别。尤其是当我们事先就期望看到因果联系时，这种偏差经常会被忽略，就像在"自尊"的例子中那样。充满诱惑的相关性证据加上固有的偏见，就能够欺骗最聪明的头脑。下面让我们看一些事例。

从关于"美国教育质量"的全国性讨论中可以很容易地看到选择性因素的重要性，这场讨论已经在美国全国范围内持续了近二十年。在这场辩论中，公众被各种教育统计数据所淹没，但研究者却没有提醒公众，警告他们避免从相关数据去推论因果关系，因为相关数据内含有大量具有误导性的选择性偏差。

纵观这场辩论，许多怀有政治目的的人试图不断地提出证据，用以说明教育质量和教师的薪资水平、班级规模是没有关系的，尽管已有许多研究表明这二者都非常重要（Ehrenberg，Brewer，Gamoran，& Williams，2001；Finn & Achilles，1999）。他们所提到的证据当中，有一个是 50 个州的 SAT（学术能力评估测试）的成绩。这个测试的参加者是有意升入大学的高中生，测试中的学生平均分确实表明，学生成绩和教师薪资水平、教育的支出是没有关系的。即使有关系，其趋势看起来也与期望的方向相反。在许多州，教师薪资水平很高，但是 SAT 的测试成绩很低，有些州教师的薪资水平在全国工资水平排行垫底，而学生的 SAT 测试成绩却很高。对这组数据的仔细审视给我们上了另外一课：

选择偏差导致虚假相关是多么容易。

举例来说，在进一步的检验中，密西西比州学生在 SAT 考试中的得分高于加利福尼亚州学生（Powell & Steelman，1996；Taube & Linden，1989），而且差异是非常显著的，密西西比州比加州的平均分要高出 100 分。而密西西比州的教师薪资水平在全国是最低的，这无疑会让那些鼓吹削减教师工资的人们弹冠相庆。但是，请等一下！密西西比州的学校真的好于加利福尼亚州？前者的教育水平真的高于后者？当然不是。几乎任何一个客观的指标都显示，加利福尼亚州的学校更好（Powell & Steelman，1996）。但是如果这是真的，那么 SAT 的成绩又该如何解释？

这个问题的答案要用选择性偏差来解释。SAT 和学校通常选择的那些标准化考试不同，在标准化考试中，所有学生一律都要参加。但 SAT 并不是所有的高中生都参加的，因而存在选择性偏差（Hauser，1998；Powell & Steelman，1996；Taube & Linden，1989；Wainer，1989）。只有那些希望进入大学的学生参加这个考试。这个因素就能够解释州与州之间的平均分为何存在差异，同时解释了为什么一些州有最好的教育体制，在 SAT 考试中的平均分却很低。

选择性因素在两个方面操纵了 SAT 的得分。首先，一些州立大学需要 ACT（美国大学考试）的成绩，而不是 SAT 分数。所以这些州中，只有那些打算去州外读大学的学生才会参加 SAT 考试。比起那些平均水平的学生，这些学生中的大部分最有可能拥有更好的家庭条件或者更高的学术才能。这种情况也发生在密西西比州和加利福尼亚州的考试中。密西西比州仅有 4% 的高中生参加 SAT，然而加利福尼亚州却高达 47%（Powell & Steelman，1996）。

第二个选择性因素则更加微妙。在那些教育质量高的州里，许多学生在高中毕业后，更倾向于继续接受教育。在这些州，参加 SAT 考试的学生比例高，这其中也包括一些学习成绩较差的学生。而在那些有着高辍学率、低教育质量的州中，想继续接受大学教育的学生比例很低。在这些州中，最终参加 SAT 考试的学生代表的是这些州中学习成绩比

较好的那些人。因此，他们的平均成绩自然要高于那些大部分人都参加升学考试的州。

关于 SAT 分数的这个例子也为我们提供了一个反面教材，那就是：公众如果缺乏本书所教授的简单方法论和统计思维技能，想纠正那些误导性的数据是何等地困难。在写于 1983 年的本书的第一版中，我就将 SAT 分数的错误使用源于选择性偏差这个例子收录进来。在十多年以后，也就是 1994 年的第四版中，我谈到了一篇文章，这篇文章是印第安纳州的教授布赖恩·鲍威尔（Brian Powell，1993）写的，文章分析了由政治专栏作家乔治·威尔（George Will）在 1993 年所写的一篇专栏文章，你可以猜测一下文章的内容……威尔反对公共教育支出，因为在 SAT 测试中取得高分的州，并没有高的教育支出。鲍威尔（1993）指出，威尔挑出的那些 SAT 分数特别高的州——爱荷华州、北达科他州、南达科他州、犹他州和明尼苏达州——参加 SAT 考试的学生比率分别为 5%、6%、7%、4% 和 10%，然而在美国参加 SAT 考试的总比率是 40% 以上。原因就是，在以上这些州中，要想进入公立学校，必须参加 ACT 考试，只有那些计划去州外有名望的私立学校读书的学生才参加 SAT 考试（Powell，1993，p. 352）。与之相反，在威尔列举的新泽西州，SAT 分数很低，教育支出却很高，其中有 76% 的高中生参加了这个考试。显然，相比新泽西州，在南、北达科他州参加 SAT 考试的学生配称得上是一支“精锐之师”。

在《教育研究者》（*Educational Researcher*）杂志中，心理测量学家霍华德·维纳（Howard Wainer，1993）分析了一篇文章，此文是美国传统基金会所做的一个研究，发表在 1993 年 6 月 22 日的《华尔街日报》上。这个基金会是一个带有强烈意识形态倾向的智库，一贯反对教育投入——你猜是为什么呢？——SAT 分数低的州往往是教育投入较高的。维纳的文章不仅揭示出选择性偏差引发的这种后果，还证明了如果测试使用有代表性的样本，而不是自我选择的样本来分析（National Assessment for Educational Progress，NAEP），这种关系就会呈现相反的情况：教育花费高的州有高的 SAT 成绩。

鲍威尔和斯蒂尔曼（Powell & Steelman，1996）利用先前提到过的偏相关技术证实了这种关系。他们发现，一旦在统计上控制了每个州参加考试的学生在比例上的差异，那么每个学生的教育支出每增加1000美元，这个州的SAT考试平均成绩就会提高15%。尽管大多数的证据都无可辩驳地表明，如果不进行统计校正，选择性偏差会让州与州之间SAT成绩的比较变得毫无意义，但媒体和政客们仍在继续使用未校正的分数来达到他们的政治目的。

来自临床心理学的例子可以表明，选择性偏差问题是多么地具有欺骗性和违背常理。研究数据有时会显示，接受心理治疗的人在各种成瘾症——如肥胖、吸毒、吸烟——的治愈率方面，要低于那些没有接受过心理治疗的人（Rzewnicki & Forgays，1987；Schachter，1982）。你想知道原因吗？原因并不是因为心理疗法使得成瘾的行为更加难以改变，而是因为那些寻求心理治疗的人的成瘾问题更复杂和棘手，而且很少能够自愈。

维纳（Wainer，1999）给我们讲了一个二战期间的故事，这个故事提醒我们选择性偏差违背常理的一面。他提到一位飞机分析师，这个分析师一直试图通过分析飞机被子弹击中的弹孔分布，来确定飞机上的哪个部位是应该放置加固防弹层的位置。他最后的决定是：把加固防弹层放在返航机上没有弹孔的地方。他的理由是，子弹袭击飞机各个部位的几率是均等的，所以，如果一架飞机能返回，就表示这架飞机被子弹击中的地方必定不会对飞机造成致命损伤。那些没有弹孔的地方，看来都是要害，因为该部位如果被击中，飞机可能就不会返航。因此加固防弹层应该安装在返航机没有被击中的部位！

总之，这一章提供给读者的规则很简单：提防选择性偏差的发生；当只有相关时，应避免因果推论。不可否认，复杂的相关数据里确实存在着有限的因果关系。同样不可否认的是，相关的证据有助于证明假设的聚合效度（见第8章）。然而对于心理学知识的消费者来说，宁可站在怀疑的角度，也不要被那些错误地暗示了因果关系的相关所蒙蔽。

小 结

本章的主旨是想传达这样一个理念，两个变量之间仅仅存在相关，并不能保证一个变量的变化就会导致另一个的变化，也就是说，相关并不意味着因果关系。在第三变量问题里，两个变量之间的相关并不意味着它们之间存在直接因果路径，因为相关的产生可能是由于这两个变量或许都与未被测量的第三变量有关。事实上，如果潜在的第三变量也经过了测量，就可以用相关统计，如偏相关（第 8 章将会讨论）来评估第三变量是否决定了这种关系。让相关统计的解释变得困难的另外一个原因，就是方向性问题。实际上，如果两个变量有直接的因果关系，因果关系的方向是不能根据相关来判断的。

在行为科学中，选择性偏差是造成诸多虚假相关的罪魁祸首。事实上人们在一定程度上选择他们的环境，并人为创造了行为特性和环境变量之间的相关。正如戈德伯格的例子所阐述的那样（在接下来的两章中我们将会进一步讨论），确保选择性偏差不会捣乱的唯一方法，就是在操纵所有变量的情况下进行真正的实验。

6

让一切置于控制之下
——聪明汉斯的故事

这一章开始前，咱们先来做一个小测验。噢，别担心，不是考你前几章所学的内容。问题其实很简单，是有关现实世界中常见的物体运动方面的知识，问题只有三个。

首先，你需要一张纸。想象如下场景：一个人拿着一根细绳在他的头顶上绕圈，绳子的另一端系着一个球。画一个圆来代表从上方俯瞰这个球的运动轨迹。在这个圈的一处画一个点，然后用一条线把这个点和此圆的圆心连接起来。这条线就代表那根细绳，那个点就代表特定时刻的球。想象在某一旋转瞬间，细绳断了。你的第一项任务是用笔画出这个球飞出后的运行轨迹。

第二个问题，假设你是一个轰炸机的飞行员，现在正以每小时 500 英里的速度在 20000 英尺（约 6096 米——译者注）的高空飞向目标，为了简单起见，假设没有空气阻力。问题是：什么地方是投掷炸弹的最佳位置，是在到达目标地点之前，还是目标的正上方，或者是在你经过目标之后？无论你选择的是目标之前、目标正上方，还是飞越了目标之后，都请你指出投放点与目标的具体距离。

最后，想象你正拿着一把来复枪从肩膀高度开火，假设没有空气阻

力，且步枪与地面是平行的。如果子弹从与枪相同的高度落地需要 1.5 秒钟的时间，那么假设你现在由枪管中射出一发子弹，初速度是每秒 2000 英尺（约 609.6 米——译者注），那么子弹落地需要多长时间？

答案——对了，还有答案这档子事儿。答案会在本章的后面揭晓。但在此之前，为了便于理解掌握这些运动方面的知识与心理学有什么关系，我们需要先深入地探讨实验逻辑的本质，这些实验逻辑经常被科学家们所使用。在本章，我们将要讨论实验控制和操纵的一些原理。

斯诺与霍乱

在前一章我们讲到，约瑟夫·戈德伯格对糙皮病的研究，一定程度上是受"糙皮病是不会传染的"这种预感的指引。但是比戈德伯格早 70 年，约翰·斯诺（John Snow）在对霍乱起因的研究过程中，则将病因放在相反的猜想上，但同样获得了成功（Goldstein& Goldstein，1987；Tufte，1977）。早在 19 世纪 50 年代的伦敦，人们对不断暴发的霍乱提出了许多理论，并且彼此争论不休。很多医生认为霍乱病人呼出的气体会将此疾病传染给别人，此理论被称为"秽气理论"。但是，斯诺却提出，该疾病是通过被病人排泄物污染的供水系统传播出去的。

斯诺开始着手验证他的理论。幸运的是，当时伦敦有许多不同的供水源，每个供水源给不同的地区供水，所以不同供水系统受感染的程度不同，霍乱的发生率应该因供水源受污染程度的不同而存在差别。但是斯诺发现，这种比较会出现严重的选择性偏差（请回想一下第 5 章的讨论）。在伦敦，不同地区的贫富差距非常大，因此，供水系统和各地区患病率之间的任何相关都会受到其他能够影响健康的、与该地区的经济发展水平相关的变量的影响，如饮食、压力、工作危机或生活质量。简而言之，获得虚假相关的可能性很大，这和第 5 章所讨论的糙皮病和污水的关系类似。但是斯诺非常机敏地注意到了一种已经出现过的特殊条件，并利用这一点解决了问题。

在伦敦的一个市区，碰巧有两家自来水公司对同一个社区供水，但从供水布局上来说是杂乱无章、毫无规划的。在某条街道上，一部分住宅是由其中一家自来水公司供水，一部分是由另外一家自来水公司负责供水，这种情况发生的原因是由于最初两家公司存在竞争。甚至有这样的情况，一栋房子由一家公司供水，而与它毗邻的房子却是由另一家公司供水。因此斯诺找到了几个由两家公司分别供水的家庭，并且这些家庭的社会经济地位基本相同，或至少是非常接近的。如果两家自来水公司都受到污染，那么这种选择仍旧是没有任何意义的，因为这样斯诺就不能发现水污染与霍乱的发病率有什么关系了。所幸的是，这种情况并没有发生，这两家公司的水并未同时受到污染。

在一波霍乱流行过后，兰姆博斯（Lambeth）公司为了避免水污染，将公司迁到泰晤士河的上游，而南沃克 - 沃克斯霍尔（Southwark & Vauxhall）公司却仍然固守在下游。因此，兰姆博斯公司的水系统受污染的可能性比南沃克 - 沃克斯霍尔公司要小得多。斯诺通过化学检验也证明了这一点。剩下的工作就是统计由两家不同公司供水的家庭的霍乱发病率：兰姆博斯公司供水的每 10000 个家庭里有 37 人死亡，南沃克 - 沃克斯霍尔公司供水的每 10000 个家庭里有 315 人死亡。

在这一章我们要讨论的是，斯诺和戈德伯格的故事是如何体现科学思维的逻辑性的。如果不能理解这种逻辑性，科学家们的所作所为看上去就会显得很神秘、怪异或是荒唐透顶。

比较、控制和操纵

尽管市面上关于科学方法论的书已经汗牛充栋，但是对于从未做过实验的外行人士来说，这些书可能都如同浮云一般，因为外行人只想知道一个大概，并不想搞清楚实验设计的所有复杂细节。科学思维最重要的特点很容易掌握，那就是科学思维所基于的理念是比较、控制和操纵。要想获得对一个现象更为深入的了解，科学家就要比较世界上存在的各

种情况。没有这种比较，我们所观察到的都是一些孤立的事件，并且对这些孤立的观察结果也解释不清，就像我们第 4 章所讨论的见证叙述和个案研究一样。

科学家通过比较在不同条件下得到的结果，可以排除一些错误的解释，并证实正确的解释。通过比较，科学家可以排除许多先前被当作既定答案的理论。也就是说，他们试图在一个实验设计中尽可能多地排除错误解释。他们这么做的途径无外乎两种：要么是在实验条件下直接进行控制；要么在自然情境下观察，以便比较各种可能的解释。

后一种情形在霍乱这个例子中得到了很好的诠释。斯诺并不是简单地随意选择两家自来水公司，他清楚自来水公司可能给不同地区供水，并且这些地区的社会经济水平会有很大差异，这种社会经济水平的差异很有可能会影响人们的健康水平。仅仅观察不同地区霍乱的发病率，难以避免"同时存在许多不同解释"的问题。斯诺清楚地知道，科学的不断发展需要尽量减少对一个问题的各种不同解释（请回想一下第 2 章所讨论的可证伪性），因此他不断寻找并且最终找到一种比较方式，此方式可以排除一大堆解释，这类解释都是与健康有关的社会经济地位方面的因素。

斯诺幸运地找到了一种自然情境，这种情境使得他能够排除其他的可能性。这种在自然情况下产生的"比较"条件并不多见。让科学家坐在那里等待这类情况发生是十分荒谬的。事实上正相反，很多科学家都试图以一种区分各种不同假设的方式来重构世界。为实现这一目的，他们必须操纵被认为是诱因的变量（在斯诺的实验里是被污染的供水系统），然后在保持其他所有相关变量不变的情况下，观察是否会有不同的结果（霍乱的发病率）。被操纵的变量称为自变量，随着自变量变化而变化的变量称为因变量。

因此，一个好的实验设计应该是这样的：科学家能够操纵他感兴趣的变量，并对其他可能影响实验的无关变量进行控制。需要注意的是，斯诺并没有这么做。他不可能操纵供水系统的污染程度，但是他找到了这样一种条件，即供水系统受污染的程度是不同的，并且与社会经济水

平有关的其他变量侥幸得到了控制。可是这种自然发生的情境不仅很少见，而且也不如直接的实验操纵那么有说服力。

约瑟夫·戈德伯格就是直接操纵变量，他假设这个变量就是引起某种特别现象的原因。戈德伯格不仅对与糙皮病相关的变量进行观察和记录，他还在一系列研究中直接操纵了其他两个变量。回想一下，他安排了低蛋白饮食的囚犯组来诱发糙皮病，同时安排吞食糙皮病患者排泄物的志愿者，其中还包括他妻子和他自己。因此，戈德伯格不仅观察了自然发生的情境，还创设了特殊条件组，从而排除一系列其他可能性并获得实验结果，这种推论要比斯诺的方法更具说服力。这也正是为什么科学家要试图操纵一个变量并保持其他所有的变量不变的原因：为了排除其他的可能性。

随机分配与操纵共同定义了真实验

我们这里并不是说斯诺的方法毫无可取之处。但科学家们的确愿意更为直接地操纵实验变量，因为直接操纵变量能够产生更具说服力的推论。细想斯诺的两组被试：一组由兰姆博斯公司供水，另一组由南沃克-沃克斯霍尔公司供水。由于处在同一个地区，可能保证了两组被试的社会地位几乎相同。但是类似斯诺这类实验设计的缺陷是：它是由被试决定自己属于哪一个组的。因为他们早在几年前已与两家自来水公司签订了供水合同。我们还必须考虑为什么一些人与这家公司签约，而另外一些人与那家公司签约。是不是一家公司比另外一家公司的口碑好？是由于这家的价钱比较便宜，还是广告说这家的水有很好的药用价值？我们不得而知。关键的问题是：这些人选择其中一家公司是不是因为该公司做广告说他们的产品质量优于另外一家，特别是对人的健康有益处？而或许这些因素才是低发病率的真正原因。这是有可能的。

类似斯诺这样的实验设计就无法排除那些更为微妙的虚假相关，这类虚假相关不像其他与社会经济地位有关的相关那样容易被看出来。这就是科学家倾向于直接操纵他们感兴趣的变量的原因。当操纵变量与一

种叫做随机分配的程序（在随机分配中被试不能决定自己进入哪种实验条件，而是被随机分配到某一个实验组）相结合时，科学家们就能够排除那些可以归因为被试本身特征的解释了。随机分配确保被试在对比实验条件下的所有变量基本保持一致，随着样本数量的增加，它还能平衡掉一些偶然因素。这是因为被试的分配是由不带偏见的随机方法实施的，而不是由某个人的选择决定的。请注意这里的随机分配与随机样本不是一回事，这两者的区别我们将会在第 7 章进行讨论。

随机分配是一种将被试分配到实验组和控制组的方法，以保证每个被试有同样的几率被分到其中一个组。掷硬币就是一种决定某一被试分到哪一组的手段。实际实验中往往采用电脑生成的随机数字表。通过使用随机分配，研究者在研究之前就试图平衡两组的所有行为变量和生理变量，甚至是那些研究者没有进行专门测量或考虑到的变量。

随机分配的效果如何，取决于实验中被试的数量。也许你会认为被试越多越好，也就是说，分配到实验组和控制组的被试的数量越多，两组间除了自变量以外的其他所有变量就越接近。但幸运的是，对于研究者来说，其实每组只需要一个相当少的人数（例如 15~20 人），随机分配就可以起到很好的效果。

使用随机分配能有效避免由于分组方式所导致的系统误差。这两组被试在所有变量上均得到匹配，但即使存在一定程度的不匹配，随机分配也消除了实验组或控制组之间的偏差。如果我们了解一下“重复”这个概念，对于随机分配如何去除系统误差这个问题就比较好理解了，所谓的重复是指在各种环境下重复一个实验，看还能否得到同样的实验结果。

设想一下，一个发展心理学家想要做一个关于早期丰富体验对学前儿童的影响的实验，在日托期间，随机分配到实验组的儿童每天接触心理学家设计的大量丰富活动，随机分配到控制组的儿童在同样的时间里只是参加一些比较传统的游戏活动。因变量是儿童上学一年后的期末成绩，通过成绩考察实验组儿童的表现是否优于控制组儿童。

像这样的实验就会用到随机分配，以确保两组在实验之初，所有能

够影响因变量的无关变量都基本保持一致。这些无关变量有时被称为干扰变量。这个实验中的干扰变量可能会是儿童的智力测验成绩和他们的家庭环境。随机分配将会在大体上使两组间在这些变量上保持平衡。但也有例外，尤其当被试人数很少时，两组仍然有可能存在差异。例如，如果随机分配之后，实验组儿童的智力测验的成绩是 105.6，控制组的是 101.9（尽管恰当地使用了随机分配，这种差异还是有可能发生），我们就会担心实验组的学业成就的任何变化缘于该组儿童的智力测验成绩高，而不是由于他们经受了丰富的体验。这里就能看出重复验证的重要性了。后续研究进行随机分配之后，两组仍然可能存在智商差异，但是随机分配程序避免了系统误差，这就能够保证这种差异不会总是出现在实验组。事实上，无系统误差这一点所确保的是，在一定数量的类似研究中，智商差异出现在实验组和出现在控制组的概率是相等的。在第 8 章我们将会讨论如何使用这种多重的实验来提高结论的聚合效度。

因此，随机分配程序有两个优点。一个是在任何实验中，样本的数量越大，随机分配越能平衡两组所有其他的无关变量。而即使在一些匹配得不是特别好的实验里，由于随机分配克服了系统误差，仍然可以让我们得出令人信服的结论——只要研究可以被重复。所以，尽管随机分配不能保证被试在任何实验中都保持完全的匹配，但它仍然是能够用来确保达到某种平衡的最好方法。

控制组的重要性

科学研究中不乏由于缺乏真实验的完全控制而得出错误结论的例子。罗斯和尼斯贝特（Ross & Nisbett，1991）提到一个发生在 20 世纪 60 年代中期的案例：门腔静脉分流术一度是一种非常流行的治疗肝硬化的方法。1966 年人们开始对此疗法进行大量研究，并且发现了一种令人感兴趣的现象。在 96.9% 的不包含控制组的研究中，医生判断这种治疗方法的效果至少在中等程度以上。在有控制组但没有使用随机分配的研究中（因此不属于真实验设计），86.7% 的研究显示同样的结论。但是，

在有随机分配的控制组的研究中，只有25%的研究显示同样的结论。因此在今天，这种特殊治疗方法被认为是无效的，但在当时，由于没有进行完全的实验控制，治疗效果被夸大了。罗斯和尼斯贝特（1991）指出，“没有使用较为正式的实验程序所获得的积极效果，要么是‘安慰剂效应’的产物，要么是由于没有使用随机分配而产生的偏差”（p. 207）。罗斯和尼斯贝特还继续探讨了“当没有使用随机分配的时候，选择性偏差是如何产生虚假相关的”这一问题。例如，如果一些病人被选作某种治疗方法的研究被试，他们可能会努力做一名好的参与者，或者他们拥有家庭的支持、积极的态度或者他们的家人对其病情更为关心，这些都可能影响实验组与控制组的差别，而这与治疗方法的效果没有任何关系。

在下结论之前必须获得“比较信息”，这种思维倾向并不是与生俱来的，这就是为什么所有科学研究都要经过训练。这些训练包括强调控制组的重要性的研究方法课程。控制组和实验组很像，只不过缺少一种重要因素的影响。控制组的这种“非鲜明性”很难让人发现它的重要性。心理学家们做了大量的研究来说明人们为什么忽视重要的比较（控制组）信息。例如，在一个研究范式中（Fiedler & Freytag，2004；Novick & Cheng，2004；Stanovich & West，1998），我们给被试呈现一个2×2的实验数据矩阵：

	好转	没有好转
接受治疗	200	75
未接受治疗	50	15

表格中的数字代表每种情况的人数。具体来说，200人在接受了治疗后表现出病情好转，75人接受治疗但没有任何好转，50人没有接受治疗但仍有好转，15人没有接受治疗也没有任何好转。研究者让看过这一矩阵的被试指出治疗是否有效，很多被试认为测试中的治疗方法是有效的，相当多的被试甚至认为治疗是很有效的。这是因为他们首先关注的是200个接受了治疗且好转的那一组，其次，他们关注这样

一个事实，即接受治疗且好转的人数（200）要远远多于没有好转的人数（75）。

事实上，这个实验所检测的疗法是完全无效的。为了理解为什么这个疗法是无效的，有必要关注一下表示没有接受治疗的控制组（没有接受特殊疗法的组）的两格数据。我们可以看出，控制组的 65 人中有 50 个人，即 76.9% 的人即使没有接受特殊治疗还是有所好转。这与 275 中 200 人（72.7%）接受治疗且有所好转形成了对比。因此，控制组中病情好转者的比例实际上更大，这说明这种疗法是完全没有效果的。只关注实验组的结果而忽视控制组的结果，会诱使许多人认为这种疗法有效。简而言之，它很容易让人们忽略这一事实，即当我们对治疗效果进行解释时，控制组的结果是背景信息中极为关键的一环。

聪明汉斯——神马的故事

用实验控制来排除某种现象的各种不同解释，这么做是极为必要的。这种必要性可以通过行为科学中一个非常著名的故事来说明。故事的主人公叫聪明汉斯（Clever Hans）——一匹会算术的马。80 多年前，一名德国教师向大家展示了一匹马，它的名字叫聪明汉斯，它好像知道如何算术。训练员无论给汉斯出加法、减法还是乘法题，汉斯都能用它的蹄子敲出答案，并且它的回答完全正确。

许多人对于聪明汉斯的表现都感到惊讶和迷惑。难道这匹马真的证明人们低估了这个物种的实际能力吗？人们无疑会有这样的疑问。对汉斯特殊能力的有力见证被德国媒体广泛报道。柏林的一家报社记者写道："这匹会思考的马将会使科学家对很多问题做很长时间的思考"（Fernald, 1984, p. 30），这个预言后来被证明是正确的——尽管与记者所期望的有所不同。一组"专家"对汉斯进行了观察，并且证明了它的能力。因此每个人对此都感到很困惑。这个困惑一直困扰人们，因为这个现象总是被孤立地观察到，也没有进行任何的控制。但这个谜团很快被一位叫奥斯卡·芬斯特（Oskar Pfungst）的心理学家解开了，他对汉斯的能力

进行了系统的研究（Spitz，1997）。

芬斯特继承了实验设计的优良传统，系统地对动物表演的环境进行操纵，创设了一种“人为”情境（见第 7 章），这种情境可以用来检验关于马的表现的各种不同说法。在一系列小心谨慎的测试之后，芬斯特发现，这匹马的确具有一种特殊能力，但不是计算能力。事实上，这匹马更像是一位行为科学家，而不是数学家。你看，汉斯是一个非常细心的人类行为的观察者，当它正在敲出答案的时候，它会观察训练员或者出题者的头部。当汉斯接近答案的时候，训练员会下意识地稍微歪一下他的头，然后汉斯就会停下来。芬斯特发现这匹马对视觉线索极其敏感，它能察觉头部的细微动作。于是芬斯特想出了另外一个方法来测试马的能力：就是让不知道答案的提问者向这匹马提问，或者让提问者在马的视线范围以外呈现问题，而在这些情况下，汉斯就失去了它的“数学能力”。

汉斯的例子很好地揭示了仔细区分“对现象的描述”和“对现象的解释”是何等重要。这匹马能够正确敲出训练员呈现给它的数学问题的答案，这是毋庸置疑的，训练员也没有撒谎，而且许多观察者也都证明了这匹马能够做到这一点。问题出现在下一步：即推论这匹马能敲出正确答案是因为它具有数学能力。推断马具有数学能力只是这一现象的一种“假设的解释”。从“马能敲出正确答案”就得出“马具有数学能力”的结论是不符合逻辑的。别忘了，马具有数学能力只是针对马的表现的诸多解释中的一种，而这种解释是可以通过实证方法来检验的。当放在这样一种实验情境下，这个解释就被证伪了。

在芬斯特涉足此事之前，那些见过这匹马的专家们都犯了一个根本性的错误：他们没有想到，对于马的表现还可能存在其他的解释。这些专家认为，只要证明训练员没有撒谎，并且这匹马真的能敲出正确答案，就能够推论出这匹马具有数学能力。然而芬斯特想得更科学一些，他意识到这只不过是众多可能性中的一种，有必要设立控制条件来区分这些可能性。于是芬斯特设计了一个情境，让训练员站在隔板的后面把问题呈现给这匹马，通过这种方式，芬斯特就可以对两种可能性进行区分：

是这匹马真的具有数学能力，还是它能对视觉线索做出反应？如果这匹马真的具有数学能力，让训练员站在隔板后面就不会对马的表现产生任何影响。而如果这匹马是对视觉线索做出反应，那么就会影响马的表现。当后者出现的时候，芬斯特就能够排除“这匹马具有数学能力”这种错误的解释（Splitz，1997）。

这里可以同第3章中讨论过的节省原则联系起来，所谓的节省原则就是说，当两种理论拥有同样的解释效力时，我们倾向于选择那个比较简单的理论（涉及较少的概念和概念之间的关系）。此处有两种理论，一种认为这匹马具有数学能力，另一种则认为这匹马是在辨别行为线索，这两种理论在节省原则上的差异是很大的。后者不需要对先前任何心理学和大脑方面的理论做出大幅度调整，它只需要我们将“马对行为线索具有敏感性”的看法稍加调整即可（现在已经广为人知）。而前一种认为马真的能学习算术的理论，则需要我们修改进化论、认知科学、比较心理学和脑科学中的很多概念。这可是相当麻烦的，因为它与其他这些科学缺乏一致性，因此如果它是真的，就需要我们更改这些科学中的很多概念才行（我们将会在第8章讨论所谓的关联原则）。

20世纪90年代的聪明汉斯

聪明汉斯的故事只是一个历史案例，很多年来，在研究方法课上，这个例子都被用来说明实验控制的必要性。没有人认为聪明汉斯的事情会再次发生，但却真的发生了。在20世纪90年代初，全世界的研究者们都在惊恐中观望，就像用慢镜头的方式观察一场车祸一样，眼看着现代版的聪明汉斯的悲剧又一次展现在他们眼前（Gardner，2001；Hines，2003；Twachtman-Cullen，1997）。

自闭症是一种严重的发展性障碍，其表现是社交缺陷、语言发展的滞后及异常，以及活动和兴趣范围狭窄等等（Frith，2003）。许多自闭症患儿从外表看起来都很正常，只是极度缺乏与人的交流，这让家长们很难接受。因此20世纪80年代末期和90年代初期，在澳大利亚有

人发明了一种技术，能让自闭的孩子从不说话到自由交流，很难想象这些自闭症患儿的家长们听到这个消息时该是多么激动。这种能让自闭症患者与人交流的技术被称为“辅助沟通疗法”，被一些很有知名度的媒体如《60 分钟》、《大观》杂志和《华盛顿邮报》等拿来大肆宣扬（见 Dillon，1993；Jacobson，Mulick，& Schwartz，1995，1996；Mulick，Jacobson，& Kobe，1993；Twachtman-Cullen，1997），据此技术的发明者称，自闭症患者以及其他因发展不良导致言语缺失的儿童，只要把手和胳膊放在这台善解人意的“辅助器”上，就可以在其辅助下，在键盘上敲出相当有文采的句子来。自闭的孩子从之前有限的语言行为到能够交流表达，这种惊人的表现无疑给沮丧的家长们带来了无限希望。这个发明者还宣称，这种技术对于那些有严重智力障碍的失语儿童也同样有效。

尽管家长们的激动心情是可以理解的，但专业人员的轻信盲从就让人不能原谅了。更为糟糕的是，在没有进行控制实验的研究之前，这些媒体节目就开始向抱有无限期望的家长们大肆宣扬这种辅助沟通疗法多么有效。要是这些专业人员在实验控制原则方面受过哪怕一丁点儿训练，他们就能立刻看出这不过是“聪明汉斯”事件的翻版。那些辅助器可以说是一个永远关注孩子成功的、富有同情心的“人”，在辅助过程中有许多机会有意或无意地指导孩子触碰键盘上的按键。另外一项观察发现，孩子们有时即使不看键盘也能打出复杂的信息，这说明辅助器给了孩子某种暗示。甚至连没学过字母的孩子也能用英语创作出优美的散文。

许多有控制的研究报告称，他们通过适当的实验控制检验了这种辅助沟通疗法。每项研究都明确地说明了同样一件事：自闭症患儿的表现依赖于辅助器发出的不易被觉察的提示（Burgess，Kirsch，Shane，Niederauer，Graham，& Bacon，1998；Cummins & Prior，1992；Hudson，Melita，& Arnold，1993；Jacobson et al.，1995，1996；Jacobson，Foxx，& Mulick，2004；Mostert，2001；Shane，1993；Spitz，1997；Wegner，Fuller，& Sparrow，2003）。在这些研究中使用

的控制方法与聪明汉斯的经典案例是相似的。研究人员设置了一种实验情境，给孩子和辅助器各自呈现一个物体的图案，但是他们彼此看不到呈现给对方的图案是什么。当孩子和辅助器看到的是相同图案的时候，孩子能正确地打出图案的名字；但是当孩子和辅助器看到的图案不同时，孩子打出的是辅助器看到的图案的名字，而不是孩子自己看到的那个图案。因此，答案是由辅助器而不是孩子决定的。实验结论是，辅助沟通疗法只不过是一种“聪明汉斯”现象，绝非治疗方法上的重大突破，也没有给研究人员带来任何欣喜。心理学家罗伯特·康明斯（Robert Cummins）与马戈特·普莱尔（Margot Prior）总结道，“这证明了一些助理人员通过利用触觉 / 视觉线索或是自身的动作姿态等来操纵客户的反应。目前研究证据所给我们的结论是令人不快的，也是不可逆转的”（p. 240）。

但悲剧后面紧跟着更大的悲剧。在一些治疗中心，有当事人在接受辅助器帮助的沟通过程中，讲出过去他们曾受到父亲或母亲的性虐待（Dillon，1993；Seligmann & Chideya，1992；Spitz，1997；Twachtman-Cullen，1997）。于是这些孩子们被迫从家里搬出来，直到这场指控被证明是毫无根据之后，孩子们才被接回来：“许多法院的诉讼案件，皆因接受辅助沟通疗法所引起，其中最引人注目的是关于控告虐待罪名的案例。法院下令的查证都证实是辅助器不适当地影响了交流，这些指控就都被撤销了（Smith & Belcher，1993，p. 176）。”哈德森（Hudson，1993）等人报告了一个实验测试：一个和父母住在一起的 29 岁女子被认为有严重的智力障碍。在接受辅助治疗期间，这个女子报告说曾经受到生活中重要人物的性侵犯。在进入法律程序时，哈德森等人通过使用芬斯特在聪明汉斯案例中用到的逻辑方法，在接受辅助疗法治疗时对这个女人的交流能力进行了测试。通过两个不同的耳机，分别对这个女人和辅助器进行提问。当两者被问到相同问题的时候，这个女人每次都能回答正确，当问到不同问题的时候，这个女人的答案都是错的，有 40% 的答案是在回答辅助器被问到的问题。

由于这些研究结果，专家的意见终于穿透媒体的喧闹浮出水面。重

要的是，大家越发认识到，这些缺乏实证基础的疗法并非无害（“哦，它有作用，那么它要是没有作用呢？”），将未经证实的疗法投入使用是要付出代价的。波士顿儿童医院的沟通促进中心的负责人霍华德·施恩（Howard Shane，1993）直截了当地指出：

> 所有科学证据都表明，辅助沟通疗法是没有用的……由于这种方法存在潜在的伤害性，所以不应被继续使用下去了。譬如，辅助沟通会导致错误的性虐待指控、人们对教育资源配置的不当要求，并且大量的训练和研究经费被投入到无用的技术里。（p. 11）

俄亥俄州立大学儿科及心理学教授詹姆斯·姆里克（见 Mulick, Jacobson, & Kobe，1993）更是一针见血地指出了这种教育手段风行一时所付出的代价：

> 如果没有对辅助沟通疗法的大力宣传，我们可能就会把更多的人力和金钱用于发展基于更有实证基础的、更可行的长远策略，来解决困扰儿童的这一问题。辅助沟通疗法的支持者为研究和专业文献所带来的理论上的混乱，对能力缺陷及其成因方面知识的积累造成了极大的损害……将辅助沟通疗法与其他成功治愈残疾人的非语言交流系统混为一谈，会使真正有效的方法也失去公众的支持……根据我们的经验，残疾人能够成为他们家庭和社区里有价值的成员，他们无需求助于神奇的治疗方法。他们可以寻求现有的有效帮助，这种帮助是有科学意义的。受过科学训练且富有同情心的专业人员的努力胜过所有流行的治疗方法，而且始终如此。治疗的进步和对于治疗的理解是建立在严格的训练、精确的科学标准以及对各种治疗理论的客观证明之上的。（pp. 278–279）

上述这个例子再次证明，仅仅相信见证叙述或者认为流行的治疗方法和伪科学无害，最终会带来危害（见第 4 章）。由此我们还能发现，当我们想要正确解释某种行为的时候，实验控制和操纵是不可替代的。

这里需要再次强调一下节省原则。自闭症儿童严重的语言障碍居然

能够通过一种“神奇子弹”式（见第 9 章）的干预方法得到治愈，而这种干预方法推翻了几十年来关于自闭症儿童的认知、神经心理和脑特征的研究成果（Baron-Cohen，2005；Frith，2003；Leslie 2000；Sigman & Ruskin，1999；Wellman，Cross，& Watson，2001）。这需要我们修改很多关于认知和神经科学方面已取得的知识。辅助沟通疗法的现状表明，它与其他科学研究成果没有关联性和一致性（见第 8 章）。

对变量分开考察：特殊条件

戈德伯格与糙皮病的例子给我们上了重要的一课，对于我们澄清有关科学进步的一些错误概念有很大的帮助，尤其是当其运用到心理学中的时候。世界上发生的任何事情通常都与其他许多因素有关联。为了对许多同时发生的事件所造成的因果影响分别进行考察，我们必须创设一些通常情况下不会出现的条件。科学实验将世界上原有的相关分割开来，以此来使单一变量的影响显现出来。

心理学家采取的也是同样的方法：通过操纵和控制来分离变量。例如，认知心理学家们对阅读的过程很感兴趣，他们对促进或阻碍文字识别的因素进行了研究。毫无疑问，他们发现较长的单词比较短的单词更难识别。乍一看，我们会认为单词长度的影响是很容易测量的：简单地设置两组单词，一组长的，一组短的，然后测量两组读者识别速度的差异。不幸的是，事情远没有那么简单。长度较长的词，其使用频率可能也较低，而使用频率本身也会影响识别。因此，长词与短词之间的任何差别都可能是由于长度、使用频率或两个因素共同作用而造成的。为了明确到底词的长度能否独立地对词的识别造成影响，研究者必须创造一些特殊的词，它们的长度与使用频率不是同时变化的。

与之类似，戈德伯格之所以能够做出强有力的原因推断，是由于他设置了一组非自然发生的特殊条件（想一下他的一个实验操纵是要被试吃下人体的排泄物，这是何等的“不自然”啊！）。回想一下奥斯卡·芬斯特设置的一些测试“聪明汉斯”的实验条件，其中包括一些提问者也

不知道答案。那些仅仅观察马在自然条件下（提问者知道答案）回答问题的人，非但永远不可能发现那匹马是如何做到这一切的，反而会得出错误的结论，认为那匹马真的具有数学知识。

同样，在检验“辅助沟通疗法”的疗效时，研究者也必须设计一些特殊的条件。呈现给辅助器和儿童的刺激必须分离，这样任何一方都不知道呈现给对方的刺激是什么。为了测试某种现象的不同假设，这类不同寻常的条件是很必要的。

心理学上的很多经典实验都需要将现实世界的自然关系分开考察，通过这样一种逻辑，就能看出哪个变量是决定因素。心理学家哈里·哈洛（Harry Harlow）的著名实验（Anderson & Anderson，1996；Harlow，1958）就是个很好的例子。哈洛想要测试一种关于亲子依恋的假设：依恋的产生是由于母亲为婴儿提供食物。然而，问题是母亲提供的不仅仅是食物（还有舒适、温暖、爱抚以及刺激等）。哈洛创设了一种条件，在这种条件下只有一个变量与依恋有关——他让刚出生的短尾猴只能在“人造的”母亲之间选择，并测查了小猴子在这种条件下的行为。例如，他发现，小猴子喜欢厚绒布做成的“母亲”所提供的接触舒适感，甚于喜欢铁丝网做成的“母亲”。出生两周之后，小猴子更喜欢冰冷的厚绒布“母亲”，而不是温暖的铁丝“母亲”，这说明接触上的舒适感比温暖更吸引小猴子（Harlow & Suomi，1970）。最后，哈洛还发现，即使当食物仅来自于铁丝“母亲”的时候，小猴子仍然更喜欢厚绒布母亲。因此，“依恋仅是由于母亲提供食物”的这种假设是错误的。正是因为哈洛能够对现实世界里同时发生的变量分开进行考察，才会有这样的发现。

创设特殊条件来验证是否存在真正的因果关系，这种方法可以防止错误观念像病毒一样侵袭我们（Dawkin，1993；Distin，2005；Stanovich，2004）。让我们看一下关于治疗性触摸的案例，治疗性触摸是在20世纪90年代北美地区十分流行的一种护理方式。使用治疗性触摸法的医生按摩的不是病人的身体，而是病人身上所谓的“能量区”。也就是说，医生的手在病人身体上方游移，但不做真正的按摩。医生说

这是在“感觉”病人的能量区。你会发现，这种感应能量区的能力可以通过创设类似于“聪明汉斯”和“辅助沟通疗法”中的特殊条件来进行验证。也就是说，测试这些医生在看不见的情况下，是否还能感觉出他们的手正接近人的身体。研究结果与聪明汉斯和辅助沟通疗法的案例一样，当视线被挡住之后，这种对距离的感觉能力和随机水平差不多。

简而言之，科学家们用创设特殊条件的方法来验证某种现象的假设是十分必要的。仅观察自然情境还远远不够，人们对下落的和移动的物体观察了几个世纪，却没有人得出关于运动和重力的正确原理和规律。直到伽利略和其他科学家们通过创设人工的条件来观察物体的运动之后，才得到了正确的运动规律。在伽利略的时代，几乎没有人看到过光滑的铜球从光滑的斜面上滚下来。世界上有很多运动发生，但这种运动却非常罕见。这是一种非常规的情境，和其他类似情境一样，使我们第一次得出运动和重力的定律。说到运动定律，在本章最开始的时候，你不是做过一个小测验吗？

直觉物理学

本章开头出现的三个问题实际上是引自约翰·霍普金斯大学的心理学家迈克尔·麦克科劳斯基（Michael McCloskey）的一本书。麦克科劳斯基研究的主题被他自己称之为“直觉物理学”。所谓直觉物理学就是普通人对物体运动的观念。有趣的是，这些观念通常与物体运动的实际情况恰恰相反（Catrambone，Jones，Jonides，& Seifert，1995；Riener，Proffitt，& Salthouse，2005）。

例如，第一个问题里，当细绳被剪断后，小球会向与细绳垂直的方向直着飞出去（即圆的切线）。麦克科劳斯基发现三分之一的大学生都回答错了，他们认为小球会沿抛物线飞出去。当麦克科劳斯基的被试被问到类似于轰炸机飞行员的那个问题时，有大约一半的人认为应在目标的正上方投掷炸弹，这就表现出他们不理解物体的初始运动决定其后来的运动轨迹，实际上应该在飞机到达目标之前五英里的地方投弹。被试

的错误不是因为问题的抽象性质所导致的。当要求被试从房间的一头走到另一头，在走的时候把一个高尔夫球丢在地板上的一个位置时，超过一半人的表现说明，他们不知道高尔夫球下落的时候还会继续向前运动。最后一道题，许多人不知道从步枪射出的子弹落地的时间与子弹垂直落到地面的时间是相同的。

你可以算一下自己在这个小测验中的成绩如何。如果最近你没有学习物理课的话，那么你很有可能至少会错一道题。“物理课！”你可能会提出抗议，“我最近当然没上过物理课，这个测验不公平！”但是请等一下，你为什么需要上物理课才知道这些题目的答案呢？从小到大，你肯定无数次地见过下落的物体。你看到过它们在自然情境中下落的过程。每天你都能看见运动的物体，你看到的是它们“自然发生”的状态。你当然不能说你对于物体运动毫无经验。当然，你没见过类似子弹的这种运动。但是我们中的大多数人都见过孩子放开旋转的物体，并且多数人也都见过物体从飞机上落下来。此外，很难说你没见过这些真实的情境。既然你有这么多年关于物体运动和下落的经验，当和真实情境略有不同的时候，为什么你不能准确地预测会发生什么呢？

麦克科劳斯基的工作很好地说明，理解科学家这一做法有多么重要。尽管人们有大量关于物体运动和下落的经验，但对于运动的直觉理论都是相当不靠谱的。我们需要明白的是，外行人观念的不准确是因为他的观察是“自然的”，而不是像科学家那样进行实验控制。因此，如果你在本章开头的测验中错了一道题，不要觉得是自己无知或知识匮乏。要知道几个世纪以前，这个世界上一些伟大人物观察下落的物体后得出的有关运动的物理知识不比现代的高中二年级的学生准确到哪去。在《科学美国人》(*Scientific American*)杂志上的一篇文章中，麦克科劳斯基指出，他观察过的被试中很多人都对物体运动持有一种错误的观念，并且这些错误的观念与在牛顿之前三个世纪的理念不谋而合。麦克科劳斯基的当代被试和中世纪哲学家有共通之处：两组人在现实世界里都有很多有关物体运动的经验，但是没有人特意创设一种条件，进行科学的操纵、控制和比较。

再多的个人经验也不足以阻止人们产生关于物理运动的错误概念。飞行员威廉·兰格威斯基在阐述关于飞行中转弯角度这一知识的发展史的时候指出，他发现飞行员在 20 世纪早期的时候拒绝使用陀螺仪这类装置，因为他们相信“平衡本能”。但是，这些“直觉”不能使飞行员在云中感觉出飞机旋转的角度。一些坠机事件和险些坠机的事件发生之后，飞行员终于清醒地认识到：没有什么直觉能够代替真实的飞行物理学知识（Langewiesche，1993）。

直觉心理学

哲学家保罗·丘奇兰德（Churchland，1988）曾指出，如果我们关于物体运动的直觉（或世俗）理论都是不准确的，因此，也很难相信我们在人类行为这类更为复杂领域中的世俗理论会是正确的：

> 我们最初关于运动的世俗理论是相当混乱的，而且最终将会被更成熟的理论完全取代。早期我们关于宇宙结构和活动的世俗理论也十分离谱，它们之所以依然存留下来，只不过是作为一些历史教训，提醒我们自己可以荒谬到什么程度。我们关于火的本质、生命本质的世俗理论也都是十分荒唐的。由于我们大部分的世俗理论都被推翻了，所以你可以一直列举下去……但是与刚才列出的内容相比，人类的心智活动是一种更复杂和难以理解的现象。目前为止才算有了一些准确的认识，而当我们在其他方面都犯了错误的时候，想要在一开始就能正确地认识心理学知识，简直就是天方夜谭（p. 46）。

生物学家 E. O. 威尔逊（E. O. Wilson，1998）道出了丘奇兰德的推测为什么可能是正确的，原因在于：“大脑这部机器构造出来是用于生存的，而不是用来理解它自己的。因为这两个目的完全不同，大脑在没有接受科学知识时观看这个世界是零散的，它必须把注意力放在它所必须知道的世界的一隅，以便于能够生存下去。这也就是为什么甚至到今

天人类了解汽车远胜于了解他们自己的大脑。”（pp. 96–97）。

当我们审视有关人类行为理论的文献时，会发现丘奇兰德的思考是对的。在第 1 章，我们证明了有关人类行为的许多常识是错误的，这不过是个小的例证而已。例如，没有证据显示有宗教信仰的人比没有宗教信仰的人更无私（Paloutzian，1983；Smith，Wheeler，& Diener，1975）。许多研究显示，笃信宗教的程度与参加慈善活动、帮助贫困的人或是不欺骗其他人这些行为之间没有直接关系。并且，在很多研究文献中，没有证据证明很虔诚的人比那些自认为是无神论者的人更慈善，或是更愿意帮助别人。

假如列出所有错误的世俗观念，那么这个清单会很长。例如，很多人认为“月亮盈亏会影响人的行为”，其实并非如此（见 Byrnes & Kelly，1992；Culver，Rotton，& Kelly，1988；Rotton & Kelly，1985）。一些人认为“性格互补的人相互吸引”，他们也错了（见 Buss，1985；Buss & Barnes，1986）。一些人认为“亲生厌，熟生蔑”，实际没有这回事（见 Bornstein，1989）。一些人认为盲人幸运地拥有超灵敏的听觉，但事实上他们没有（见 Niemeyer & Starlinger，1981；Stankov & Spilsbury，1978）。类似这样的例子还有很多很多。

人类关于行为的直觉理论是有缺陷的，这就说明了为什么我们的心理学研究需要实验控制。只有这样，我们才能把我们关于人类行为的粗浅概念上升为准确的科学概念和体系。

小　结

实验方法的核心就是操纵与控制。在相关研究中，研究者仅仅观察两个变量的自然变动是否显示某种联系，而在真实验中，研究者要对被假设为原因的变量进行操纵，通过实验控制和随机分配来保持其他所有变量不变，然后来看这个假设变量是否会产生影响。这种方法排除了相关研究中出现的第三变量的问题。第三变量出现的原因是，在自然情境

下，很多不同的事物都是相互联系的。实验方法就是用来分开考察这些自然存在的关联。它之所以能实现这一目的，是因为它以操纵一个变量（被假设是原因的变量）的方式分离出该变量，并保持其他所有变量不变。但是，为了区分这些自然的关联，科学家们经常要创设自然世界里不会出现的特殊条件。

7

“但是这不是真实的生活！”

——“人为性”批评和心理学

前两章讲述了实验逻辑的原则，现在我们可以思考一下心理学经常面对的一些批评。比如很多人认为科学实验没有价值，因为它是人为发生的，和“真实的生活”不一样。我们将对这一观点进行详细探讨。由于心理学实验常常遭到类似的批评，因此理解这种批评的不合理之处，将有助于我们更好地了解心理学。

为什么自然性并非总是必要的

从第 6 章的内容中，我们已经可以清楚地看到为什么这种批评是不合理的。正如上一章所述，科学实验的人为性并不是一种缺点，事实上，正是它使得科学方法具备了一种奇特的力量，可以让我们对世界进行解释。与人们通常所相信的观点不同，科学实验的人为性并不是偶然的疏忽，而是科学家故意为之。科学家之所以专门设置一些非自然发生的条件，是因为只有这样才可以将决定事件发生的许多相关变量区分开来。

有时候，必要条件已经在自然状态中存在，比如斯诺和霍乱病的

例子。但这种情况并不经常出现。科学家必须用新异的甚至有时比较奇怪的方法操控事件，比如戈德伯格和糙皮病的例子。很多时候，这些操作无法在自然环境中完成，于是科学家必须把所要研究的现象转移到实验室中，以便实施更精确的控制。例如在有关“重力和运动”的早期研究中，使用了一些特制的物体，其目的就是为了创造一些特殊条件，以便观察物体运动。因此，为了分析一种现象，经常需要创设非自然的极端条件。

事实上，如果科学家完全禁锢在“自然”条件下观察，那么一些现象就不可能被发现。探索物质本质特征的物理学家们建造巨大的加速器来诱发基本粒子之间的碰撞。碰撞中产生的一些副产物是存在时间不到十亿分之一秒的新粒子。然而，这些新粒子的属性却有助于解释原子结构理论。许多新粒子在世界上一般是不存在的，即使存在，我们在自然状况下也没有机会观察到它们。因此，几乎没有人质疑物理学家们的研究方式。为了对宇宙有更深刻的理解，即使采用一些不常见的甚至是怪异的方法，也是合情合理的。但不知为什么，物理学家用起来合理的方法，心理学家使用起来，就常被认为是不合理的。

由于公众没有意识到创造特殊条件的重要性，各类科学家都曾遭到过误解，而心理学家是这种误解的最大受害者。许多心理学家在向外行人展示关于某一行为的实验证据之后，都听到过这样的叹息：“可惜这不是真实的生活”。对这种批评的进一步讨论，通常可以让我们了解到大众的一些观念，比如，只有研究自然条件才能获得知识，心理学的实验室研究是怪异的，还不足以成为一门科学。

心理学家使用的许多技术在公众看来是怪异的，很多人都不知道这些技术并非心理学领域所独有，只不过心理学家把这些科学方法应用到人类行为的研究上而已。事实上，在每一个学科的调查研究中，都能看到人们用一些类似的怪异方式来获取关于世界的知识。心理学家却遭受了双面夹击。对科学方法的无知使得许多人相信心理学永远不可能成为科学，于是，当心理学家像其他所有的科学家一样创造必要的特殊条件来对心理现象进行更有力、更精确的解释时，却遭到了诽谤。

禁锢于真实生活条件会妨碍我们发现许多新事物。例如，生物反馈技术现在被广泛应用于各种领域，比如用于控制周期性偏头痛和紧张性头痛、治疗高血压，以及放松训练（deCharms et al.，2005；Maizels，2005；Miller，1985a）。研究表明，如果通过视觉或听觉的反馈能够监测到体内正在进行的生理过程，那么人类就能学会在一定程度上控制这些过程。这项研究促进了上述生物反馈技术的发展。当然，因为人类本身并不具备通过外部反馈来监测自身生理功能的能力，所以，如果不是在特殊的实验室条件下，人们将很难发现人类有能力控制自己的生理过程。自然条件下的观察是永远无法发现这一点的。

对“随机取样”的误解

然而有时候，类似“这不是真实的生活”的抱怨源于对心理学实验研究目的的另一种误解，产生这种误解的原因是非常容易理解的。媒体的宣传使许多人对调查研究开始熟悉起来，特别是选举中的民意调查。现在人们对选举投票的一些重要特征越来越了解。具体而言，为了保证民意测验的准确性，媒体对随机取样、样本代表性等概念更加关注。这种关注导致许多人错误地认为，随机取样和代表性是所有心理学调查研究的必要条件。因为心理学研究很少使用随机的被试样本，如果根据外行人所相信的随机取样标准，那么许多心理学的研究成果都会遭到诋毁，那些批评心理学研究无法反映真实生活因而是无效的论点也会受到强化。

但只要想一下其他科学的情况，就很容易理解这种想法的荒谬。化学家从没尝试过抽取化合物的随机样本，生物学家也不曾用细胞或组织的随机样本进行实验。用于医学研究的老鼠和猴子也不能完全代表其物种。而这些研究都是在与这些动物生活的自然环境完全不同的实验室中进行的。事实上，这些条件通常很独特。然而，所有这些研究得到的结果都可以帮助我们理解人体生物学。大部分心理学研究也是同样的道理。并非每一个心理学调查研究都需要使用随机样本。因此，我们在此需要

强调的重点是：随机取样和随机分配（见第 6 章）不是一回事情。

随机分配和随机样本的区别

随机分配和随机取样两个词里都包含“随机”，因此许多人以为它们所指的是一回事。事实上，它们是非常不同的概念，唯一相似之处在于它们都采用了随机生成数字这一点。然而其目的却大相径庭。

随机取样涉及的是如何选择被试进行研究。如前所述，并不是所有研究都要求随机取样，但当它成为必要条件时（例如在调查研究、市场调查或是选举时的民意调查中），我们则需要用一种方法从总体中抽取一个样本，这种方法要确保总体中的每一个成员都有同等机会被选为样本，被抽中的样本就成为随后调查研究中的被试。有一点非常重要，这种随机抽样的调查研究既可能是相关研究，也可能是一个真实验。只有使用了随机分配的方式，才有可能成为一个真实验。

随机分配是真实验所必需的条件。实验人员将被试分为实验组和控制组，当每一名被试被分到实验组的机会和被分到控制组的机会相等时，则实现了随机分配。为了达到这一点，常会用到像掷硬币这样的随机化手段（更常用的是一种特殊的随机化数字表格）——因为它在给被试分组时没有任何偏向。

随机分配和随机取样不是一回事，牢记这一点的最好方法是弄清楚四种组合：非随机分配的非随机样本，随机分配的非随机样本，非随机分配的随机样本，以及随机分配的随机样本。大部分心理学实验没有使用随机样本，因为没有这个必要。正如下一章将讲到的，研究可以检验理论，我们所需要的只是一个方便取得的样本。如果一个研究中使用了随机分配的方法，那么它就是一项真实验，如果没有使用，那么它是一项相关调查。许多使用随机取样的研究没有使用随机分配，那是因为它们只是调查性研究，旨在寻找关联——也就是说，这些研究属于相关调查研究。然而，一些研究既使用了随机取样，又使用了随机分配，那么它们肯定是真实验。

理论研究和应用研究的异同

弗吉尼亚大学心理学家道格拉斯·穆克（Douglas Mook）阐述了不同类型的研究要求的不同类型的预测。许多应用研究的目的是把研究结果直接与生活中的特殊情境联系起来。对应用研究而言，预测必须是：研究与现实生活有“一对一”，亦即穆克称作“类比”的关系，应用研究的结果必须都能直接应用。选举投票中的民意测验就是应用研究的一个例子。研究目的是预测一个特定情境下的特定行为，在这个例子中，就是选举日的投票结果。由于研究结果是要直接应用于现实的，因此，样本的随机性和情境的代表性问题很重要。

然而，把应用型心理学研究看做典型的心理学研究是错误的。心理学（或其他学科，就这一点来说也是如此）的大部分研究都有着不同于应用的目的。大多数研究的预测层次是从理论到特定研究情境的预测。大多数研究的结果只能间接通过理论修改而被应用，这些理论与其他科学规律共同应用于一些实践性问题（Nickerson，1999）。简而言之，大部分理论研究追求的是对心理过程的理论验证，而不是把研究结果推广到现实中的某一特殊情境中去。

主要目的为理论验证的研究通常被称为“*基础研究*”。应用研究的目的是把数据直接应用于现实生活，但是基础研究则专注于理论验证。然而，仅仅根据某项研究是否有实践性应用来区分基础研究和应用研究，很可能会产生错误，因为这一差别常常会随着时间的增长而逐渐消失。应用研究的结果会很快得到应用。但是没有什么能比普遍的、准确的理论更具有实用性了。尽管很多科学家进行理论或实证研究的初衷并非解决具体的实践性问题，但他们发展出的科学理论或研究结果最终都解决了现实世界的许多问题。这样的例子在科学史上不胜枚举。

罗伯特·克里斯和尼古拉斯·萨米奥斯（Robert Crease & Nicholas Samios，1991）在一个处于领先地位的物理学研究中心——布鲁克海文国家实验室（Brookhaven National Laboratory）工作，他们在一篇论文中介绍了一些有关基础研究拥有应用价值的例子。比如威尔赫姆·伦琴

（Wilhelm Roentgen）的故事，他“发现了一件奇妙的事情，当他把荧光屏放在他的仪器附近时，荧光屏就会意外地发光；结果发现了自然界的一个新现象，他称之为X射线。三个月之后，X射线已经被人们用来检查骨折”（p. 82）。再来看霍华德·弗洛瑞（Howard Florey），他和一个同事一起研究抗菌机制。“青霉素是他们所研究的微生物之一，这种微生物是几年前才偶然发现的，但它的抗菌功能一直没有被发现。现在的许多药物都是在一些对应用没有特别兴趣的学术研究发现的基础上研制出来的。”（p. 82）

西莫尔·凯迪（Seymour Kety，1974）也谈到了氯丙嗪这一治疗精神分裂症的特效药物是如何在一些看似没有关联的科学研究中被发现的。凯迪强调，几乎所有应用于精神分裂症治疗方法的发现都与精神分裂症毫无关系！历史一再证明，（通过让科学家解决特殊的实践性问题而）试图控制科学发展方向只能阻碍发展进程而非促进。具有讽刺意味的是，急于让科学家们解决实际问题，而不让其考虑“其他事情”（基础研究）的做法，被证明是最不切实际和目光短浅的。

通向实际应用的道路是非常曲折和不可预料的。为了研究关节炎，德克萨斯西南大学医药研究中心的一组研究人员试图通过遗传的方式培养一批患有关节炎的老鼠。出乎意料的是，这些老鼠同时也患上了肠炎（Fackelman，1996）。这个研究团队“偶然创造出了有溃疡性肠炎的老鼠，从而为科学家们研究人类疾病提供了动物模型”（Fackelman，1996，p. 302）。无论这些科学家是否在关节炎（原本想研究的问题）上取得了进展，现在看来他们似乎在溃疡性肠炎的治疗上做出了巨大的贡献。

简言之，我们必须意识到，虽然一些研究是为了直接预测某一特殊情境而设计的，但大多数科学研究仍然是用于验证理论的基础研究。怎样把研究结果应用到现实生活中呢？从事应用研究和从事基础研究的研究者们对此有不同的回答。前者会这样回答：“直接应用，只要实验情境和将来要应用的情境有相当程度的相似性就可以了。”因此，被试的随机取样和实验情境的代表性都会影响结果的应用。然而，进行理论检验的研究人员是这样认为的：研究结果不会直接应用于现实生活，进行

理论研究的目的也不是为了将结果用于具体的环境条件。因此，这类科学家并不关心研究的被试与其他群体有多相似，也不关心实验情境是否反映出某些真实生活的环境。那么，这是否意味着这些研究结果对现实世界没有意义呢？不是的。这些研究结果不直接应用于某一特殊情境，而是应用于理论。这种理论，也许在将来的某一天，可以和其他科学规律相结合，共同解决某一特殊问题。

在心理学的一些领域里，这种将理论间接应用于现实生活的研究十分常见。例如，许多年前手机刚刚面世，许多认知心理学家立即开始担心安全问题——人们边开车边接听手机怎么办。心理学家立即预测手机的使用可能会导致交通事故增多——不仅仅是因为接听电话的时候一只手会离开方向盘，此外，他们还担心接听电话会转移司机的注意力。有一点很重要，我们应该意识到，心理学家提出这些担忧远远早于真正用移动电话来做的实验研究（见 Strayer & Johnston，2001）。心理学家通过理论预测手机事故问题，而这个例子中的注意力有限加工理论早在几十年前就已经存在了（如，Broadbent，1958；Kahneman，1973）。这一信息加工理论是通过大量的实验验证（上百个实验室研究）建立起来的，开车使用手机提供了一个机会，正好可以用这一理论来预测其可能造成的危害。事实也是如此，后来使用移动电话进行研究，结果证实了心理学中注意理论的预测：移动电话的使用确实是引发交通事故的一个原因（Insurance Institute for Highway Safety，2005；McEvoy et al.，2005；Redelmeier & Tibshirani，1997，2001）。

道格拉斯·穆克（Douglas Mook，1983）就一个例子阐述了心理学中通过实验来验证理论的观点以及间接应用的性质。20 世纪 30 年代，塞里格·海奇特（Selig Hecht）在《普通实验心理学手册》（*Handbook of General Experimental Psychology*）（Murchison，1934）里发表了一系列有关视觉敏感度的研究，谈到了暗适应的现象。你可能有过暂时性“失明”的经历，比如当你走进一个漆黑的电影院时。但是，当你在位置上坐了一会儿之后，应该就能注意到椅子、人以及其他物体慢慢变得可以看见。如果你继续关注这个现象，你会发现视敏度不断升高的这个过程会持续

几分钟之久。

这种现象叫做暗适应，它会经历两个阶段：首先是在进入一间漆黑的屋子时，视敏度迅速小幅度地升高，之后缓慢大幅度升高。海奇特把两部分的升高曲线和视网膜上的两种感光细胞联系起来。密集分布在中央窝中心（视网膜的一部分，用于聚光）的视锥细胞，对红光非常敏感。分布在中央窝外围的视杆细胞，没有那么密集，而且对红光不是很敏感。海奇特根据这些事实建立了一个理论，即暗适应的最初阶段（视敏度小幅地快速升高）源于视锥细胞的适应，第二阶段（在更长的一段时间内视敏度大幅升高）源于视杆细胞的适应。

穆克（1983）提醒我们考虑一下海奇特的实验环境是完全非自然的。（非随机取样的）被试在暗室里进行反应，根据他们是否察觉到微弱的红色闪光，回答“是，我看得见”或者“不，我看不见”。正常情况下，我们不会在日常生活中对微弱的红光作“是”或“否”的反应。然而由于海奇特并不考虑将自己的研究成果推广到那些在暗室里对红光做“是”或“否”的反应的个体中去，所以现实生活中这种情况是否真的发生过无关紧要。海奇特所关心的是，如何根据实验室中建立的事实来验证相应的理论，从而能解释视觉系统所特有的一些基本过程，如暗适应。他并不关心他的实验情境是否符合现实世界的情况，而是关注是否能充分分离出他想研究的特殊视觉过程。

海奇特的研究发现之所以具有普遍性，并不是因为他的实验情境的性质是人工的或是自然的，而是因为根据这些研究结果可以建立一个有关基本视觉过程的理论，而这个理论可以与许多视觉现象相关联。他的研究揭示了人类视觉系统中各个部分之间的功能关系，而这恰恰是因为他的研究情境经过了人为的精确控制。如果这一理论模型是正确的，那么它应该能广泛地应用于各种情境，可以用来解释许多行为现象，即使所处的情境与发现这一理论的情境完全不同。换句话说，海奇特的研究结果通过对理论的影响而产生了间接的应用价值。例如，海奇特的研究结果促进了对夜盲症的治疗，改善了X射线的识别问题（Leibowitz，1996；Mook，1982）。更引人注目的是，二战期间，英国飞行员在闪电

战中等待希特勒轰炸机的夜间袭击时，戴上了红色的飞行眼镜（因为视杆细胞对红光不够敏感，可以保持暗适应；见 Mook，1982）。从在实验室里判断小红点到辨别伦敦上空危险物体的移动，这一鸿沟是由理论跨越的，而不是通过把海奇特的实验室改造成喷气式战斗机得出的。

心理学理论的应用

一旦我们明白了大部分研究的目的是发展理论而不是预测具体环境下的事件，以及大部分研究的结果是通过理论间接应用的，而非在具体环境条件下直接应用，那么我们就会顺理成章地发问：究竟心理学中有多少理论可以在现实中得到应用。也就是说，心理学理论的普遍性得到验证了没有？

对于这一点，我们必须承认以往的记录是参差不齐的。但也必须清楚地意识到，这与心理学的多样性息息相关。一些领域中的研究确实在应用方面进展甚微，然而，其他一些领域则已经取得了十分瞩目的成绩，通过实验已经推导出了许多具有解释能力以及预测效力的原理（见 Buss，2003；Byrnes，2001；Davidson & Sternberg，2003；Gazzaniga，2004；Hilton，2003；Sunstein，2002；Wilson & Keil，1999；Zimbardo，2004）。例如，在咨询、教育、临床、心理治疗等应用领域，理论导向的基础研究结果也获得了大量实际应用（Davidow & Levinson，1993；Garb & Schramke，1996；Lilienfeld，Lynn，& Lohr，2003；Lilienfeld，Wood，& Garb，2000；Mahrer，2000；Siegert & Ward，2002）。

想想经典条件反射和操作性条件反射原理。这些原理及其详细论述的规律，几乎完全是从非人类被试的实验发展而来的，比如鸽子、老鼠，其实验情境也是高度人为化的实验室环境。然而，这些原理已经成功地用于解决人类各式各样的问题，包括自闭症儿童的治疗、大量事实材料的教学、酗酒和肥胖症的治疗、精神病院的病号管理以及恐惧症的治疗

等等。这仅是一小部分而已。

这些应用所依靠的原理之所以能够被准确地提炼出来，是因为在实验室条件下，研究者们能够精确地细化环境刺激和行为之间的关系，而这一点在自然条件下是无法做到的，因为在自然情境下，许多行为之间的关系可能会同时起作用。至于非人类被试的使用，是因为在许多案例中，从动物的反应得出的理论和规律为我们提供了与人类行为非常相近的数据（Vazire & Gosling，2003）。人类研究发现，人类的行为规律与从动物行为得出的规律非常相似。当人类疾病治疗方面的每一次医学进步都源自动物研究数据的时候，这些发现也不应该再让我们感到惊奇了。举例来说，动物研究促进了很多领域的发展，包括行为医学、压力缓解、心理治疗、受伤或者残疾人士的康复、衰老对记忆力的影响、帮助人们克服神经性肌肉紊乱的方法、药物对胎儿发育的影响、药物滥用、失忆、交通安全、慢性疼痛的治疗（Domjan & Purdy，1995；Gosling，2001；Kalat，1995；Miller，1985b；Zimbardo，2004）。最近，关于狗的研究为理解人类焦虑障碍的基本原理起到了实质性的推动作用（Groopman，1999）。

事实上，“这不是真实的生活”的批评被错误地用来诋毁动物研究的成果——这种做法经常是由于政治的缘故。例如，那些为重度污染企业效劳的政客们总是否认致癌因素风险评估报告的有效性，他们的理由是，这些报告是以动物研究为基础的，不能应用到人类风险评估上。然而，一组科学家在1988年进行的一个对23种致癌物质（苯、石棉，等等）的研究中发现，由动物研究计算出来的死亡率与由人类流行病学研究计算的结果非常相近（Finkel，1996）。

心理学家对知觉过程的研究取得了令人印象深刻的进展，从中得出的规律和理论已用于解决各种各样的问题，比如雷达监测系统、街灯照明以及飞机驾驶舱的设计（Nickerson，1992；Swets，Dawes，& Monahan，2000；Wickens，1992）。关于衰老对认知的影响，我们已经积累了许多新的认识（Salthouse，2004），而这些新知识有可能会直接帮助我们设计出帮助认知丧失者恢复其能力的训练方案（Dixon &

Backman，1995）。

判断和决策的心理研究已经应用于医学、教育和经济等领域的决策制定（Gigerenzer，2002；Gilovich，Griffin，& Kahneman，2002；Hilton，2003；Kahneman，2003；Kahneman & Tversky，2000；Swets，Dawes，& Monahan，2000；Tetlock，2005；Zweig，2001）。一项令人兴奋的发展是认知心理学家更多地参与到法律体系中，这使得在信息搜集、证据评估和做出判决等方面，给认知理论的应用带来了大量机会（Kassin，Tubb，Hosch，& Memon，2001；Koehler，1993；Kuhn，Weinstock，& Flaton，1994；Redding，1998；Wrightsman，1999；Zimbardo，2004）。而且从 20 世纪 80 年代中期开始，阅读教学中的理论和实践开始受到认知心理学的影响（Rayner，Foorman，Perfetti，Pesetsky，& Seidenberg，2001，2002；Snowling & Hulme，2005；Stanovich，2000；Vellutino，Fletcher，Snowling，& Scanlon，2004）。此外有关儿童在法律程序中提供的证词（Ceci & Hembrooke，1998）和受虐儿童所“恢复”的记忆是否准确（Brainerd & Reyna，2005；Bremner，Shobe，& Kihlstrom，200；Clancy，Schacter，McNally，& Pitman，2000；Gardner，2006；Loftus，1997；Lynn，Loftus，Lilienfeld，& Lock，2003；McNally，2003；Pezdek & Banks，1996；Spanos，1996）等这些公众争论不休的问题方面，心理学家都提供了重要的科学依据。美国心理学会开设了一个网站：www.psychologymatters.org。你可以在此网站上看到更多有关心理学知识的实践性应用。

“大二学生”问题

许多人质疑心理学研究成果的代表性，他们过于关注研究的被试，而不关心实验设计的细节。然而，心理学的许多领域中，研究结果更多受到后者的影响而非前者，例如知觉基本过程的研究。在基本的信息加工过程、大脑的基本结构以及视觉系统的特征方面，美国蒙大拿州的人和佛罗里达州的人都非常相似。另外，这些特征和被试的父母是不是修

补匠、裁缝或者教授没有任何关系。

所有学科都假设某些因素对最终的结果没有影响。生物学家一般不担心培养皿厚度的细微差异会显著影响里面的细菌。当然，这些差异可能会产生影响——每一种科学假设都不是绝对的——但是生物学家必须把精力放在他们认为更有可能产生影响的变量上面。与此相似，海奇特假设暗适应跟人的宗教无关，因此他不问被试是路德教派的，还是罗马天主教的。

我们面临一个问题，有时人们叫它“大二学生问题”；即，因为大二学生在大量的心理学研究中做被试，因此这些研究所得出的结果是否具有可推广性受到了质疑。心理学家关心这一问题，是因为它在某些研究领域中的确是个问题。尽管如此，我们还是要正确地看待它，并且应该知道心理学家对这一批评有几种合理的辩解。以下列出三点：

1. 这种批评不能说明研究结果无效，只是需要更多的研究来证明理论的可推广性。由于我们先前收集了大二学生的数据，即使从其他人群中获得了相反的数据，从而必须对理论做出相应的调整，也只会使理论更加精确，而不会完全否定它。即使在比较极端的情况下，重复的实验没有得出相同的结果，我们也只能说，建立在大二学生数据基础上的理论不够全面，而不能说该理论一定是错误的。
2. 在心理学众多领域里，大二学生问题不构成一个问题，因为所研究的心理过程是非常基本的过程（例如视觉系统），几乎没有人相信视觉系统的基本构造跟被试样本的人口分布特征有关。
3. 许多研究结果得到了重复，这使我们确信这些结果在很大程度上可以推广到不同的地理分布中，并且在较小程度上，也能推广到具有不同社会经济因素、家庭变量以及早期教育经历的人群中去。50年前的大学生被试样本恐怕基本来自于精英团体，如今却完全不同，现在大学生的家庭背景能够代表各阶层的群体。

然而，不承认大二学生问题在心理学研究的某些领域里的确是个问题，这也是不明智的（Peterson，2001）。尽管如此，心理学家正在尽力

矫正这个问题。例如，发展心理学家几乎都很关注这个问题。这一领域中每年都有成百上千的研究人员，将众多用大学生被试得出的理论和发现，在其他不同年龄的被试身上重新验证。

用不同年龄组的人做被试并不总是能重复用大学生被试得出的结果。要是那样的话，发展心理学就会变得很无聊了。但是一大堆心理学家都致力于在心理学理论中建立一个年龄因素，以证明这个因素的重要性，这一领域的研究也确保了心理学的宏大理论不是只建立在从大学生那里收集的有限数据基础之上。

许多发展过程的研究都是以北美儿童为被试进行的，为了评估这些研究结果的可推广性，发展心理学家也进行跨文化的研究。例如，史蒂文森等人（Stevenson et al.，1985）让中国、日本和美国孩子完成大量的认知任务，然后得出结论，“三种文化下的儿童在这些任务中所表现出的认知能力非常相似”（p. 727）。其他认知能力的跨种族和跨文化比较研究也证实了相同的结果（Demetriou et al.，2005；McBride-Chang & Kail，2002）。对于其他心理学特征，跨文化研究里有许多例子也显示了相似的趋势（例如，Day & Rounds，1998；Rozin，Lowery，Imada，& Haidt，1999）。但是，也有不少跨文化研究没有得出与美国大二学生相似的结果（例如，Nisbett，2003）。但是当出现这些差异时，这些研究仍然提供了一些重要信息，让人们了解到这些理论和结果会因文化和背景的不同而不同（Chang，1996；Menrich et al.，2004；Medin & Atran，2004；Nishett，2003；Yates，Lee，& Shinotswka，1996）。

正如先前提到的，认知心理学的研究成果通过了重复验证。信息加工的许多基本规律在全世界许多实验室中得到验证。人们可能不太知道，如果密歇根大学的一名心理学家获得一项重要的研究成果，那么类似的实验将很快在斯坦福大学、明尼苏达大学、俄亥俄州立大学、剑桥大学、耶鲁大学、多伦多大学等大学进行。通过这种检验，我们将很快知道这项结果是不是由于密歇根州被试的独特性或特殊的实验环境所造成的。

教育心理学家也提出过大二学生问题。例如，教育心理学家联合发展心理学家以及其他教育研究人员，构建了测量基本认知技能的内容，

这些内容能在一定程度上准确预测学业成绩，例如阅读习得率。这些内容的预测准确度不受社会经济地位和种族的影响，也不受地理分布和校区的影响。

大二学生问题和关于代表性的批评大部分针对的是社会心理学，社会心理学经常用大学生被试在实验室情境中进行研究，并试图建立真实社会情境中的社会交往、群体行为和信息加工等理论（Kunda，1999）。然而，即使在心理学的这一领域，也有证据表明，实验室得出的成果和理论，实际上确实预测出了不同类型的个体在各种情境下的行为。

例如，几年以前，莱昂纳德·伯克维茨（Leonard Berkowitz），威斯康辛大学的一名心理学家证明了所谓的“武器效应”——如果一件武器出现在手边，会使得某个人更容易做出攻击性反应。这个发现源于实验室，是一个无代表性情境的典型例子。由于这一结果是人为情境的诱导产物，因此常被强烈地批评其具有误导性。但事实是这样的，各种实验条件下得出的结果都一样，用不同的方法测量攻击性所得的结果一样，在欧洲和美国获得的结果一样，研究儿童和成人的结果一样，在实验室之外的现场研究中，被试不知道自己是在参与实验，得出的结果也一样（Berkowitz & Donnerstein，1982；Turner，Simons，Berkowitz，& Frodi，1977）。研究人员甚至提取出了武器效应背后的认知机制。在语义记忆中，它是一个自动启动的过程（见 Anderson，Benjamin，& Bartholow，1998）。这个效应是如此之强，以至于安德森等人（Anderson et al.，1998）把文章题目定为“是手枪自己扣动扳机的吗？”

认知、社会和临床心理学家也研究了人类的各种决策行为。这个研究领域里大部分原始的研究都是在实验室里完成的，使用大学生作被试，而且采用高度人为化的任务。然而，从这些研究中得出的决策行为原则在很多非实验室环境中都得到了重现，包括银行家对股票价格的判断、赌场赌博、精神病医生对病人行为的预测、经济市场预测、军事情报分析、全美橄榄球联赛的下注、工程师对修理时间的估计、房地产经纪人对房价的估计、商务决策以及医生的诊断——这些原则现在也应用于个人理财咨询的实践领域（Bazerman，1999，2001；Belsky & Gilovich，

1999；Goldstein & Hogarth，1997；Hammond，1996；Tassoni，1996；Zweig，2001，2002）。

伯恩鲍姆（Birnbaum，1999）用互联网来解决心理学中的大二学生问题。他在实验室里通过互联网招募了一批参与者，并进行了一系列有关决策问题的实验。实验室中得到的结果全部在互联网样本中得以重现，而后者的取样范围要比前者广泛得多——包含来自 44 个国家的 1 224 名参与者（见 Jaffe，2005；McGraw，Tew，& Williams，2000）。高斯林等人（Gosling et al.,2004）研究了大量互联网参与者的样本（361 703 人），并将之与发表过的 510 个传统样本的参与者比较，发现互联网上的被试在性别、社会经济地位、地区和年龄方面有着更广泛的分布。重要的是，他们发现，心理学众多研究领域的研究结果，例如人格理论，用互联网实验和传统方法的研究所得出的结果非常相似。

这些例子说明，心理学研究结果的稳定性和普遍性程度经常被低估了（Cheng，2001；Gage，1996；Rosenthal，1990）。安德森、林德西和布什曼（Anderson，Lindsay，& Bushman，1999）对实验室研究和现场研究的效果进行了系统检验。从验证攻击性、领导力、抑郁症等多种主题的研究中，安德森等人（1999）发现了高度的聚合性——几乎所有不同的研究环境总是产生相似的结果。

当然，并不是所有的心理学研究成果都能重现（见 Carroll & Nelson，1993）。相反，结果无法重现的实验经常出现，而它们往往比结果可以重复的实验更具指导意义。但是，在认知心理学中，重复实验的失败几乎很少是被试的独特性造成的。相反，大部分是因为实验刺激和方法的细微差异。通过仔细地分析要产生一个现象究竟需要哪些实验条件，科学家们对现象有了更精确的理解，这为建立一个更精确的理论奠定了基础。

但是，如果实验结果没有被重现，那么心理学的研究成果如何应用？如果科学家们没有在所有的细节上达成一致，知识和理论并不完全站得住脚，那么如何证明这些结果的应用是合理的呢？这种对心理学发现的担心是很常见的，因为人们没有意识到，在其他科学中，结果和理论经

常在完全确立之前就开始应用了。当然，第 2 章中已经清楚地阐述过，所有的科学理论都有可能被修订。如果我们在应用科学研究结果之前必须确定知识是完全正确的，那么应用就不会发生了。所有领域的应用型科学家尽最大努力使用最准确的信息，同时也会意识到这些信息有可能是错误的。

正确看待“真实生活”和“大二学生”问题

本章提到了几个焦点问题，此外有一点很重要，就是我们应该清楚什么是我们说过的，什么是我们没说的。我们证明了对心理学研究的频繁抱怨源自一个基本的误解，不是针对心理学，而且针对涉及所有科学的一个基本原则。我们看到人们质疑心理学家为什么不在所有研究中都使用随机样本，并且解释了这种担心是多余的。最后，我们看到，大二学生问题本来是一种合理的关注，但它有时被夸大了，尤其是当人们对心理学研究的广泛性和多样性不太熟悉时（见第 1 章）。

尽管如此，心理学家应当始终注意他们的实验结论不要太过依赖于某一种方法或某一特殊被试群体。这一点将在下章讨论。事实上，心理学的一些领域确实被大二学生问题折磨得够呛（Jaffe，2005；Peterson，2001；Wintre，North，& Sugar，2001）。作为大二学生问题的一剂良药，跨文化心理学仍然是一个亟待发展的领域。然而，研究型心理学家对于自我批评的高度重视，给了我们一个对此持乐观态度的理由（见第 12 章；Anderson & Gunderson，1991；Henriques，2003，2004，2005；Jaffe，2005；Kimble，1999；Machado，Lourenco，& Silva，2000；Proctor & Capaldi，2001）。事实上，许多心理学家因为将“批判”作为其本职工作而享誉学界（Leary，2001；Robinson，2001）。每年的各类科学杂志上，都会有文章提醒心理学者注意其方法上的漏洞，或是指出大二学生问题。后者在心理学中是一个受到广泛关注的问题，没有心理学者尚未意识到这一点。因此，尽管我们不应忽视这一问题，同时也应正确看待它。

小 结

一些心理学研究属于应用型研究，它们的目标是把研究结果直接应用于特定情境。在这样的应用研究中，研究的目的是要将结果直接推广到自然情境中，样本的随机化和条件的代表性就显得尤为重要，因为研究结果将会直接得到应用。然而，大多数心理学研究不属于这种类型，而是属于基础研究，用以验证有关行为潜在机制的理论。在大部分基础研究中，研究结果通过理论上的修正得到间接应用，从理论产生到应用于某些实践性问题需要一段时间。在这种类型的基础研究中，被试的随机取样和情境的代表性不是关键问题，因为这类研究的重点在于验证理论的普遍性。实际上，在用于验证理论的基础研究中，人为的环境条件是有意创设的，因为（正如前一章所描述的）这有助于把研究的关键变量从所要控制的无关变量中分离出来。因此，心理学实验“不像是真实的生活”这个事实其实是一种优势而非缺点。

8

避免爱因斯坦综合症
——聚合性证据的重要性

“生物学实验揭开生命的奥秘！”、“思维控制上的新突破！”、“加利福尼亚科学家发现了延缓死亡的方法！”——如你所见，想仿制一条充斥于小报和电子媒体头版头条的“突破性”新闻简直易如反掌。由于部分缺乏责任感的媒体总是定期炮制这类“头版头条”，难怪大多数科学家都建议公众要以怀疑的态度来对待此类新闻。但是，本章的目的不仅仅在于反对夸大事实、以讹传讹的做法，也不仅仅提醒人们在评估科学进展报告时必须审慎地考察其来源，我们还想提出一种比前面章节中提到的理念都更为综合、全面的科学进步观。为此，我们将会详细阐述曾在第 1 章中介绍过的系统实证主义和公共知识。

媒体上这类所谓的“突破性”头条新闻，在很大程度上误导了公众对于心理学和其他科学的认识。一个特别典型的误解就是，它们让公众以为，某一科学研究领域中的所有问题都能通过某个关键实验得到解决，或者是某一个重要的灵感成就了理论的进步，并彻底颠覆了先前众多研究者累积的全部知识。这种科学进步观非常符合新闻媒体炒作的胃口，在媒体的运作方式里，对历史的追溯就是呈现支离破碎、缺乏连贯的小型事件。对于好莱坞娱乐业来说，这也不失为一种颇为便利的模式，那

里，事件必须有一个开头和圆满的结尾，含糊的东西都被理得清清楚楚。然而，这只是对科学进步的一种歪曲，如果对此信以为真，就会导致关于科学进步的错误观念，并削弱人们在某一问题上评估科学知识的能力。在本章中，我们将会讨论科学的两个原则——关联性原则和聚合性证据原则，用这些原则描述科学发展，将比“跃进模式”更为准确。

关联性原则

在否定所有科学进步的“飞跃”或者关键实验模式的有效性的同时，我们不是说这种关键实验和理论发展模式从未发生过，相反，科学历史上一些著名案例表明这种模式的确出现过。爱因斯坦提出“相对论”就是迄今为止最著名的一个例子，至此，一系列非凡的理论灵感重新定义了时间、空间和物质等基本概念。

然而，爱因斯坦的成就如丰碑般矗立，让这种科学发展模式统治了公众的内心。这种统治是持久的，因为它与媒体报道大部分新闻事件时所采用的隐含“脚本”高度吻合。人类历史上，像相对论那样遭受了那么多的胡言乱语和不实推论的理论并不多见（不，爱因斯坦没有证明“一切都是相对的”——见 Holton，1996；Randall，2005）。当然，我们的目的不是去批驳这些谬论，而是为了后面讨论和评估心理学中的理论做铺垫。

在爱因斯坦的理论中，那些被重新定义的关于物理世界的概念是如此地基础，以至于那些通俗读物经常将其等同于艺术领域里的概念变化（一个二流诗人经过重新评估，摇身一变成了天才；一个艺术流派被断言灭亡）。这种做法忽视了概念变化在艺术和科学中最根本的差别。

科学中的概念变化遵从关联性原则，而这一原则在艺术中并不存在，或至少说是极为罕见的（见 Bronowski，1956，1977；Dobzhansky，1973）。就是说，一个新的科学理论，必须与先前已确立的实证事实建立关联。新的科学理论不仅仅要解释新的事实，还要兼容旧的事实，这

样才会被认为是一个真正的理论进步。新的理论可以以一种迥然不同的方式来解释旧的证据，但是它必须能解释得通。这些要求保证了科学在原有的基础上持续进步。除非理论解释效力的范围被拓宽了，否则真正的进步是不会发生的。如果一个新的理论解释了一些新的现象，但是没有解释大部分旧的事实，那它将不会被认为是对于旧的理论的全面超越，因此不会立即取代那些旧理论，而新理论和旧理论将会以竞争性的面貌并存，直到有一个能整合二者的新理论出现。

无论爱因斯坦理论中的那些新概念是多么地令人震惊（钟表变慢、质量会随速度增加等等），但它们都遵从关联性原则。在宣告牛顿力学的滞后性的同时，爱因斯坦的理论没有否定那些以牛顿观点为基础的运动事实，或者是认定其毫无意义。相反，在速度较低的情况下，这两种理论做出了本质上相同的预测。爱因斯坦理论的高明之处在于，它能够解释更为广泛的新现象（有时是令人吃惊的），而这些是牛顿力学所做不到的。因此，即使是爱因斯坦理论这个在科学历史上最惊人的、基础性的概念重构，也依旧遵循着关联性原则。

消费者规则：警惕关联性原则的无效性

科学发展的“跃进式”模式——我们可以称之为爱因斯坦综合症——让我们误入歧途，以为新的发现必定违反关联性原则。这一观念很危险，因为如果舍弃关联性原则，最大的受益者将是那些伪科学和伪理论的贩卖者。这些理论之所以受到青睐和关注，就是因为它们总被说成是“全新的”。“毕竟，相对论在它所在的时代是新生事物，对吧？”这句话经常被用做一种说辞，以证明某种新鲜玩意儿是正确的。当然，在这个伪科学家虎视眈眈的领域里，先前积累的事实数据看上去似乎是个巨大的绊脚石。然而事实上，这块绊脚石也无法阻挡这些伪科学家，这是因为他们有两种强有力的伎俩来化解这一麻烦。一种伎俩我们之前已经讨论过（见第 2 章），就是解释数据前先将这个理论变得不可证伪，这样就令先前的数据毫无用处了。

第二种伎俩是宣称先前的数据与他们的主题无关，因而不予考虑。为了实现“不予考虑”的结果，他们通常强调新理论呈现出“前所未有”的新颖性。类似“关于现实的全新观念”和“前所未有”这样的语句被频频使用。但实际上，真正的花招还在后面。“新理论”注定如此具有突破性，以至于源于其他理论测试的实验证据都被宣称是与之不相关的。只有能被新理论的框架所兼容的数据才会被考虑；也就是说，关联性原则被完全破坏了。显然，这个理论是如此之新，以至于他们可以理直气壮地说：与之关联的实证证据尚不存在呢。如此这般，你就拥有了一个适宜伪科学发展的优质土壤：旧的、“不相关”的数据灰飞烟灭，新的相关数据尚不存在。这种伎俩很容易得逞，因为爱因斯坦综合症蒙蔽了关联性原则。而颇具讽刺意味的是，关联性原则的重要性就是由爱因斯坦理论本身所论证的。

加利福尼亚古生物学者凯文·帕迪安（Kevin Padian）给出了另外一个例子，说明当人们无法意识到关联性原则的重要性时，会如何误解科学的本质。谈到堪萨斯州学校董事会将进化论从学生的必修课中删除的决定时，帕迪安指出：“我们在谈论有关‘科学如何被整合’的一个误解。那种认为可以抛弃一部分科学——特别是像进化论这样一个可以把整个生物学串联起来的核心理论——并认为它没有其他关联的想法，实在是太荒唐了”（Carpenter，1999，p. 117）。生物哲学家迈克尔·鲁斯（Michael Ruse，1999）指出，进化论与许多独立的科学领域都呈现出关联性，例如古生物学、胚胎学、形态学、生物地理学、神经系统科学等等。同样，赫门（Shermer，1997）指出，“如果宇宙和地球只有一万年的历史，那么宇宙学、天文学、物理学、化学、地质学、古生物学、古人类学以及关于早期人类历史的科学都将是错误的”（p. 143）。著名的科学作家和考古生物学家史蒂文·古尔德（Stephen J. Gould）赞同这一观点，“教生物学不教进化论，就像教授英语不教英语语法一样”（Wright，1999，p. 56）。

鲁斯（1999）讲述了一个例子来描述达尔文如何使用关联性原则，并舍弃了某个与其他学科之间缺乏必要关联性的新理论。当时达尔文想

探寻一种能与他的自然选择理论相匹配的遗传机制，为此他试图去建立一个所谓“泛生论”的理论。“身体的各部位都会产生一些小的胚芽，这样胚芽在体内循环，并在性器官处聚集，从而传给下一代”（p. 64）。一个问题是，这个理论和细胞学说并不一致。第二个问题是达尔文没有解释这些胚芽是怎样被运送的，因为输血试验已经证明胚芽不能通过血液传输。基于这两点以及其他一些原因，泛生论在科学阵营里被淘汰出局，“因为它与生物学的其他领域不相兼容”（p. 64）。

同样的情况也适用于心理学，哲学家马里奥·邦格（Mario Bunge，1983）已经指出，认知心理学如果一开始就否定经典条件反射和操作性条件反射，那么它将无法在心理学中立足，因为它无法兼容行为科学中的其他知识。回忆第 6 章中对“辅助沟通疗法”的讨论，它之所以不能“治疗”自闭症语言障碍，是因为它打破了关联性原则——如果治疗有效，它将会要求我们重建神经病学、遗传学和认知心理学领域内的知识。这一假设性的疗法与科学中的其他知识没有任何关联。同样，神创论对进化论的反对也是如此，神创论在科学上没有遵守任何关联性原则。相反，进化论跟其他科学有着密不可分的关联。如生物学家西恩·卡罗尔（Sean Carroll，2005）所说，“进化论不仅仅是一个生物学主题，它还是这一学科的基础。生物学没有进化论就像物理学没有重力学说一样”。（p. 52）

有这样一个来自心理学的例子。假设有两种疗法被开发出来，用于帮助有严重阅读困难的孩子缓解其问题。两种疗法都没有经过实证性的检验。第一种，疗法 A 是一个训练程序，目的是在音位水平上促进儿童对语言片段的认知。第二种，疗法 B 通过让孩子蒙上眼睛走平衡木以训练前庭器官的感受性。疗法 A 和疗法 B 在一个方面上是一致的——它们的效果都没有经过直接的实证检验，二者反响都不好。然而，其中一种疗法在关联性原则方面是占据优势的。疗法 A 与研究文献中的广泛共识具有一致性，在这些研究文献中提到，具有严重阅读困难的孩子受到阻碍，是因为孩子还没有发展出足够的对于语言片段结构的认知（Snowling & Hulme，2005；Vellutino et al.，2004）。疗法 B 没有和任何

相应的学术共识发生关联。这种关联性的差异预示疗法 A 是一个更好的选择，即使二者都还没有经过直接的检验。

“跃进”模式与渐进整合模式的比较

这种把爱因斯坦式的革新视为科学典型的倾向，诱使我们误以为所有科学进步靠的都是重大飞跃。问题就在于，人们倾向于将这些例子泛化成一种观念，认为科学进步理所应当是这样产生的。事实上，很多科学领域的进步靠的都不是某一个突然的突破，而是由一系列构不成重大影响的停顿及前进之间的反复所组成的。

科学工作的不确定性是大部分公众所意识不到的。科学实验很少能完全确定某个问题，或支持某一理论从而排除其他理论。新的理论也很少能够全面超越所有先前存在的相互竞争的概念体系。很多问题的确定并不像科学电影里所描绘的那样，由一个关键实验所决定，而是要等到科学界逐渐开始有了共识，认为支持某种理论的证据比支持其他任何理论的证据要有力得多。科学家所评估的证据不是来自于某个设计得异常完美的实验的数据，与之相反，科学家往往需要去评估来自几十篇实验论文的数据，这些实验各有瑕疵，但都能提供部分答案。科学发展的这种渐进模式受到阻碍，正是因为爱因斯坦综合症在公众中造成了一种思维定势，认为所有科学都和物理学一样，因为对于物理学来说，科学进步的跃进模式或许是最适用的。

想想遗传学和分子生物学在过去一个世纪中的突飞猛进。这些进步的产生不是因为一个爱因斯坦式的伟人在关键时刻现身，然后搞定了一切。相反，数百个存在瑕疵的实验产生出来的数十种灵感和洞见，促成了现代生物学的整合。这些进步的发生，凭借的不是革命性地重构一些重大概念，而是几种都能站得住脚的不同解释进行长期与反复的交锋和对峙。经过十几年没有定论的实验、无数次的理论构思、争辩与批判，科学家们终于弄明白了基因到底是由蛋白质还是由核酸组成的。他们达成新的共识，但不是通过一次跃进式的改变。原子核的发现者恩斯特·卢

瑟福强调了关联性原则的重要性，“科学家不能仅仅依靠一个人的观点，而是要依靠千万人的智慧”（Holton & Roller，1958，p. 166）

卢瑟福的观点强调了另一种区分科学与伪科学的方法。科学总是遵循关联性原则，其特点在于众多个体的参与，而对这些个体的贡献进行评判的标准，是看它在多大程度上加深了我们对自然界的了解。没有哪个单独的个体能够依靠其特殊地位来主导讨论。当然，在第 1 章，我们已经讨论过科学的这种公共性，相比之下，伪科学经常认为特定的权威和研究者才有接近真理的“特殊”机会。

我们曾提出过两个理念，能为理解心理学的规则提供一个有用的情境。首先，科学上没有哪个实验是被设计得完美无缺的，对任何一个实验数据的解释都存在着不确定性。科学家们评估一个理论，往往不是坐等一个完美的或者关键的实验的出现，而是对大量局部存在瑕疵的实验的总体趋势进行评估。第二，许多科学在即使没有爱因斯坦的情况下也取得了进步。这些进步是蹒跚而曲折的，而不是通过伟大的“爱因斯坦式整合”那样的阶梯式跃进。和心理学一样，其他许多科学也都是由那些原本缺乏共同主题的知识不断积累和拼接而成。

聚合性证据：在瑕疵中进步

先前的讨论引出了一个证据评估的原则，这个原则在心理学中至关重要。它常被称作*聚合性证据原则*（或者*操作聚合原则*）。科学家和那些科学知识的运用者常常不得不做出判断：海量的证据到底说明了什么。在这种情况下，聚合性证据原则就成了一个非常重要的工具。聚合性证据原则对于科学信息的外行使用者来说也是个有效的工具，尤其是在他们要对心理学主张做出评估的时候。尽管对于聚合性证据这个概念所做的详尽性技术讨论将很快让我们晕头转向，但事实上，此概念在实际应用方面的作用很容易理解。我们将探索两种方式来表述这个原则，一种是按照“瑕疵实验”的逻辑，另一种是按照理论检验。

从极端上讲，导致一个实验出错的方式有无数种（或用术语来说，就是变得混淆）。然而在大多数情况下往往不会有那么多干扰性的关键因素。在某个特定领域中拥有丰富经验的科学家，往往很清楚什么是最关键的因素。因此，当对某一研究结果进行审查时，科学家总能察觉实验中的关键瑕疵。接下来，聚合性证据原则提示我们去审查相关研究文献所呈现的瑕疵模式，因为这类模式要么支持、要么否定我们想要做出的结论。

假设来自大量不同实验的结果都很一致地支持了某一特定结论。假如实验本身并不完善，我们应该继续去评估这些研究瑕疵的性质和程度。如果所有的实验都是以同样一种方式出现瑕疵，这些情况将会降低我们对实验结论的信心，因为结论的一致性也许仅仅源于一个特定的瑕疵，而这个瑕疵是所有实验共有的；另一方面，如果所有实验都呈现出不同的瑕疵，我们对结论的信心就会大增，因为结果的一致性看似并非源自某一个让所有实验结果都混淆不清的干扰性因素。正如安德森（1996）所言，“不同的方法很有可能涉及不同的假设，当一个假设能够通过众多基于不同假设的证伪检验时，我们可以说是得到了一个强有力的结论”（p. 742）。

每一个实验都有助于纠正其他实验在设计方面的错误，而其他实验反过来也检验它的瑕疵，使其获得支持。尽管各自多少都存在不同的缺点，并且实验技术也各有优劣，但只要大量的实验能够得到近似的结果，那么我们就可以说我们的实验证据实现聚合了。即使没有一个实验设计得十全十美，我们还是得到了一个相当有说服力的结果 。因此，聚合性证据原则允许我们将结论建立在大量有些许差异的实验来源之上。这个原则之所以能让我们得出有说服力的结果，是因为这种方法所获得的结果的一致性不大可能是由某个实验程序的特殊性所造成的。

聚合性证据原则同样能以理论检验的形式加以表述。当一系列实验始终支持某个假定的理论，同时又能共同排除那些最主要的竞争性理论时，研究就具有高度的聚合性。尽管没有一个单一的实验能排除所有的可能解释，然而一系列具有一定诊断效力的实验（如果所有数据都呈现

某种特定趋势的话）就能产生一个极具说服力的结论。

例如，假设针对某一现象，有五种不同的理论（称它们为 A、B、C、D 和 E）同时存在，且都经过了一系列的实验验证。假设一部分实验以很强的效力检验了理论 A、B 和 C，结果数据否定了 A 和 B，支持了 C。再想象一下，另外一些实验则以同样的效力检验了理论 C、D 和 E，结果数据否定了 D 和 E，支持了 C。这种情况下，对于理论 C 我们就有了强有力的聚合性证据。我们不仅有了支持理论 C 的数据，还拥有了对抗其他竞争性解释的数据。强调一下，没有一个实验能够检验所有的理论，但是汇总起来，一系列实验就能做出有力的推断。

相反，如果所有已知的研究都只强有力地检验了 B、C 和 E，并且数据结果支持 C 并否定了 B 和 E，那么理论 C 的说服力就不如前面例子中那么强了。原因在于，尽管产生了支持理论 C 的数据，仍然没有强有力的证据能够排除其他可能的理论（A 和 D）。因此，当一系列实验始终支持某个假定的理论，同时又能共同排除那些非常重要的竞争性理论时，研究就具备了高度的聚合性。尽管没有一个单独的实验能够排除其他可能的解释，但如果将一系列具有部分诊断性的研究按照上述例子中的方式加以汇总，就能得到一个比较有说服力的结论。

最后，聚合性证据原则能够让我们摈弃一个误区，这个误区的形成是由于我们在第 2 章对于证伪性的讨论过分简单化所造成的。当时的讨论似乎让人觉得，当第一个与自己的理论相抵触的证据出现时，这个理论就算是被证伪了。然而事实并非如此（Pigliucci，2002）。正如理论是被聚合性证据所支持一样，它也要被聚合性的研究结果所否定。

心理学中的聚合性证据

强调聚合的重要性的原因在于，心理学结论往往是建立在聚合性证据原则之上的。这个事实当然并不独特或罕见（在其他很多的科学中，结论也不是基于单一的、决定性的实验证据，而是基于众多结果不甚明晰的实验）。但这种情况在心理学中尤为突出，心理学实验的诊断性

往往较低。也就是说，支持某一个理论的数据经常只能排除一小部分可能的解释，还遗留了许多有可能取代这种理论的"候补"理论。其结果是，只有收集并比较来自大量研究的数据之后，才能得到有说服力的结论。

心理学实验具有高度模糊性的因素，这毫不奇怪，因为其研究的问题涉及复杂的人类行为。如果心理学家坦然承认这个事实，然后耐心地去解释这个事实所带来的结果，那么公众就能够更好地理解这门科学。心理学家应该承认，尽管心理科学已然存在并且在不断进步，但这种进步是缓慢的，并且许多结论往往都来自于令人感到折磨的长时间的统合和争论之中。对于媒体经常宣称的所谓突破性进展，我们要永远带着怀疑的态度，但心理学主张所经受的怀疑却是千真万确的。

在心理学中，我们必须像走钢索一样谨小慎微。例如，我们必须抵御这样的诱惑：当证据还不确凿时，就把某一假说当作已经证实了的理论来对待。本书连续几章都反复强调了这种怀疑态度。要注意不要从相关中推论因果，拒绝接受见证叙述式的证据。与此同时，我们不应该对知识的不完整和最后结论有待探索等事情反应过度，并开始怀疑心理学究竟能否产生有说服力的结论。我们也不应该被"心理学不可能成为一门科学"这种非理性的主张所诱惑。根据这一立场，聚合性证据的原则可用来平衡对假设性知识所做的过度诠释。尽管所有的心理学研究都存在这样那样的瑕疵，但聚合性能让我们获得有说服力的结论。

证明聚合性证据原则的最好方法，就是检验心理学中一些仍存在争议的领域。让我们通过一个例子看看聚合性证据原则的重要性。这个问题是，接触暴力电视节目是否会增加孩子们的攻击性行为。对于这个问题，目前的科学共识是：观看暴力电视节目（或者电影）的确能增加儿童的攻击性行为。这种影响不是很大，但真实存在。科学家们对于这个结论的信心并非来自于某个单一的、权威的研究，而是来自许多研究结果的汇总（如，Anderson，Berkowitz，Donnerstein，Huesmann，Johnson，Linz，Malamuth，& Wartella，2003；Anderson & Dill，2000；Anderson & Huesmann，2005；Anderson，Huston，Schmitt，

Linebarger, Linebarger, & Wright, 2001 ; Bushman & Anderson, 2002 ; Paik & Comstock, 1994)。这一研究结论适用于电视游戏、电视及电影（Anderson & Bushman, 2001)。这些研究所采用的研究设计、被试规模以及特定技术都有很大差别，但现在能够清楚发现，这些差别是此领域内各种研究的优势，而不是弱点。

尽管电视所属的产业能给孩子带来负面影响，证据十分确凿，但那些电视网和电视游戏产业的老板们还是自然而然地会抵制这些证据。他们发起了一场误导公众的运动，利用的正是公众“不能意识到研究结果是建立在许多研究的聚合上而非某个单一的、具有决定意义的证明上”这一特性。电视网公司不断挑选个案并暗示，只要证明每一个研究都存在瑕疵，就可以全盘否定总体的结论。尽管社会科学研究者也许会去回应对于某个具体研究的批评，但并不能就此认为研究者总是轻易承认某个特定研究存在着缺陷。关键的区别就在于，研究者拒绝这样的暗示，即承认某个特定研究存在瑕疵，就否定了“影视暴力会对攻击性行为产生影响”这一普遍的科学共识。其原因就在于，普遍的结论来源于聚合性。即使是不包含这类瑕疵的研究，其结果也会指向同一方向。这一研究当然也有其自身的问题，但其他研究对此进行修正之后也产生了相似的结论。

例如，关于这个问题，早期研究揭示了观看暴力节目的数量和儿童攻击性行为之间的相关。这些相关证据不能被视为因果关系，指出这一点是非常正确的。也许是第三个变量导致了这种关联，也许更有攻击性的儿童选择去观看更多的暴力节目（方向性问题）。

但是科学团体的结论不单单是建立在相关证据之上的。研究者不仅对两个变量之间的关联进行简单的测量，还使用了更为复杂的相关技术，这些相关技术允许研究者得出一些因果性质的试探性结论（其中一个如偏相关，在第 5 章中曾提到）。这类技术中，有一种方法采用纵向设计，即在不同时间点测量相同的两个变量——在这里就是电视暴力和攻击性。由这一设计所得到的相关模式可以告诉我们二者是否有因果联系。有人已经进行过这类研究，得到的结果表明：观看暴力电视节目的

确有可能增加人们日后的攻击行为。

同样，有人批评纵向相关技术尚存在争议，这并非毫无道理，因为它确实有争议。关键在于，“电视暴力和攻击性行为之间存在因果联系”这一结论依靠的并不完全是简单或复杂的相关证据，因为研究者还进行了无数的实验室研究，在这些研究中，电视暴力的数目得到了直接操纵，而不仅仅是被评估。在第 6 章，我们曾讨论过变量的操纵，操纵与随机分配等控制手段共同使用，就能避免相关研究在解释问题时存在的不足。如果有两组儿童，在其他变量均得到了实验平衡之后，仍然表现出不同水平的攻击性行为；如果这两组儿童的唯一的区别就是一组观看暴力节目，而另一组没有观看暴力节目，那我们能做出正确的推断：被操纵的变量（电视暴力——自变量）导致了结果变量（攻击性行为——因变量）的变化。这个结果在大部分实验研究中都出现了。

这些研究已经激起了一些“这不是真实的生活”的非议，这种非议连同那些毫无根据的指责，在先前的章节中都曾讨论过。无论怎样，电视暴力的影响并非针对某个特定群体的儿童，因为这些结果在美国不同的地区和世界的不同国家都得到了印证。使用不同的实验情境、不同电视节目作为实验刺激的各个研究都得到高度一致的结果。

重要的是，从现场实验而非实验室实验中也得到了相同的结论。一种叫做*现场实验*的设计也被用来研究电视暴力 / 攻击性行为问题。这类研究设计的存在提醒我们，不要认为实验情境和实验设计之间存在必然的联系。有时候人们认为，我们只有在实验室里才能操纵变量，在非实验室情境中只能进行相关研究。这个想法是不正确的。实验室里也常常进行相关研究，而非实验室情境下也常常可以操纵变量。尽管有时在非实验室的情境下操纵变量进行现场实验需要相当大的创造性，但这一方法在心理学领域中被越来越普遍地采用。

当然，现场实验本身是有缺陷的，这些缺陷往往是其他研究的强项。总的来说，将观看电视暴力和儿童攻击性行为增加联系起来，所使用的证据并非仅仅依靠某一研究甚至某一类型的研究。

这种情形类似于吸烟和肺癌的关系。烟草公司的老板们经常试图去

误导公众，暗示吸烟导致肺癌的结论只基于个别研究，然后便开始批评个别研究。恰恰相反，支撑这个结论的是很多聚合性的证据。来自于不同研究的数据的聚合性是很强的，这些数据的聚合性不会因为对某个研究的批判而彻底改变。

事实上，在这里有必要讨论一个类似肺癌起因的科学问题。医学诊断和治疗中的许多决策，都建立在不同研究结果能否汇聚为一个结论的基础之上。例如，当流行病学调查（可以说是一种涉及人类的现场研究，目的在于寻求某一疾病与环境及地理因素的关联）、精确控制的动物实验以及人类被试的临床实验等不同类型的研究结果，都趋向于汇聚在一个结论上时，医学界才会对这一结论抱有较大的信心，认定这一结论是可靠的，医生们才愿意在这些证据的基础上实施治疗方案。

然而，所有这三种类型的研究都有其各自的缺陷。流行病学研究经常是相关性的，在变量之间存在虚假相关的可能性很高。实验室研究能被高度控制，但实验对象往往是动物而不是人类。医院环境下的临床试验在真正的治疗环境中使用人类作为被试，但仍有很多控制的问题，因为存在安慰剂效应和治疗病人的医疗团队的期望效应。就像吸烟和肺癌的例子中那样，尽管每一种研究都存在问题，但当来自不同方法的数据能较强地汇聚起来的时候，医学研究者们就能够做出有说服力的结论。这与心理学家可以用聚合性证据原则来帮助他们做出电视暴力对攻击性行为有影响的结论一样。

科学共识

评估电视暴力影响的问题是一个典型例子，告诉我们：在心理学中，数据最后是如何累积起来用于解决问题的。尤其是在社会急切关注的领域，切记，这些问题的答案只能在大量不同研究结果实现融合之后缓慢地出现。通过单个突破性的研究是不可能解决这些问题的。用一个简单原理来总结：在评估心理学的实证证据时，心中要想的是“科学共识”，

而不是“重大突破”；是“渐进整合”，而不是“大步飞跃”。

“共识而非突破”这一原则，可以通过有关“儿童早期补偿性教育计划”的争议加以说明。在 20 世纪 60 年代末和 70 年代初，当时关于林登·约翰逊（Lyndon B. Johnson）总统提出的“优质社会计划”是否真的有效的争论正进行得如火如荼，公众经常会在报纸上看到这样的头条：“早期干预可将智商提高 30 分”和“心智开启计划是一个败笔”等等。一个外行人该如何面对如此矛盾的信息呢？在这个例子中，“科学共识而非重大突破”原则无疑可以提供一些帮助，因为这一原则提醒我们，这两个报纸头条可能都不太成熟。事实上，研究者又花了十年的时间才在这个重要的社会问题上达成一个科学共识。

这一共识的产生并非源自于某个单独的重大研究成果，而是当康奈尔大学的一组研究人员（Lazar，Darlington，Murray，Royce，& Sniper，1982）在 20 世纪 60 年代和 70 年代早期将来自 11 个不同早期教育项目上的几百个被试的数据汇总起来分析时，这种共识才得以建立。尽管单一项目的研究结果有时候很难去解释，但当它们汇聚在一起时，整体的研究结果就非常清晰了。早期教育干预的短期项目没有顺理成章地让 IQ 增加 30 分。另一方面，心智开启计划以及一些类似项目也并没有绝对失败。早期教育干预项目的确能对参与此项目的儿童后继的教育历程产生具体的影响。这些儿童更少会留级，更少被安排到特殊教育班，而且对学校和学业成绩有更为积极的态度，并持续表现出学业成绩提高（也可见 Lee，Brooks-Gunn，Schnur，& Liaw，1990；Ramey，1999）。

加拿大心理学家提莫西·摩尔（Timothy Moore，1996）认为，如果人们能更加普遍地意识到聚合性原则，那么在法庭上将会更好地利用专家证词。他特别讨论了依赖专家证词的问题。证词属于个人意见，难以代表该领域专家的共识。摩尔引述了在“犹大圣徒”（Judas Priest，美国七八十年代著名的重金属摇滚乐队——译者注）一案中的专家证词。这起案件涉及两个青少年的自杀，他们的父母控告摇滚乐队“犹大圣徒”在歌曲中传达的潜意识信息诱发了他们孩子的自杀。尽管专家证词指出，当时的科学共识是：没有任何证据显示那些潜意识信息能产生这种效果

（即使是现在，这一共识仍然成立），然而，在一个不能反映实证性共识的学者进行了一番言之凿凿的心理动力学解释之后，这个案子的法官多少还是受了些影响。摩尔总结说，这个学者误导了法庭，“他的观点虽然极富想象力和逻辑性，但与当时对于此问题的主流科学理解相悖。长长的履历和尊贵的职位并不足以保证其观点是科学有效的，单个专家的证词是独特的、个人化的，并且未经更广泛的科学团体的认定，这样的专家不足以引导整个法庭”（p. 38）。

研究方法和聚合性原则

聚合原则同样也意味着，我们应当乐于看到多种不同方法应用于各个心理学研究领域之中。因为不同的研究技术各有其优势和不足，用于获得特定结论的各种方法之间呈现一种相对的平衡是比较理想的。心理学长期以来都因过于依赖基于实验室的实验技术而受到诟病。这种批评在心理学家之间也是个争议性的话题。然而，一种确定无疑的趋势是，近年来，心理学各个领域都已经开始使用不同的研究方法了。例如，由于过度依赖实验室技术，社会心理学家遭受的批评可能是最多的，但社会心理学家已经开始转向了更富想象力的现场设计，以寻求聚合性的证据来支持他们的理论（Kunda，1999）。

心理学者比伯·拉坦（Bibb Latane）和约翰·达利（John Darley）的研究为此提供了一个很好的例子。这些研究者们因对“旁观者现象”的研究而广为人知。旁观者现象是指，一些人在看到他人处于危难之中时并不施以援手。拉坦和达利（1970）指出了这样一个事实：在很多危急时刻，当有其他旁观者在场时，某个旁观者伸出援手的可能性会更低。

然而，这两位研究者清楚地知道，这些仅凭被试在实验室里的反应而做出的结论太过单薄了。因为在实验室中，被试都是在自愿报名到实验室来参加实验之后才目睹紧急事件的。为此，拉坦和达利设计了另外一个有趣的实验，希望在另一个情境中重现这一现象。他们找到一个愿意合作的卖酒的商店，该商店同意假装店里发生了盗窃事件。当收银员

在店铺的后面为一个“顾客”拿啤酒时，该“顾客”（实际上是研究者的同伴）拿起一箱啤酒走出店门。这一幕总发生在收银台前一个或两个真正的顾客的眼皮底下。收银员回来后问这一个或两个顾客，“嗨，刚才在这里的那个人到哪儿去了？你看见他离开了吗？”这样，就给了顾客一个机会向收银员报告刚才发生的盗窃事件。与实验室实验的结果吻合：当旁观者在场的时候，向收银员报告盗窃案的行为受到了抑制。

社会心理学家并不是唯一试图在不同的情境中重复其研究结果的人。认知心理学家们也开始探索如何推广他们的许多实验结果。例如，吉格伦泽（Gigerenzer，1984）研究了“频率-效力效应”的普遍性。这个效应是指，一个陌生但看似有理的论断，不管是真是假，只要经过不断地重复，就会增加人们对它的相信程度。这个效应成功地得到了重复验证，但是这些研究都是在实验情境中，以大学生为被试（并且绝大多数在美国）。于是吉格伦泽做了一个非实验室情境、以非大学生为被试的研究。他在德国慕尼黑测试了许多非大学生的成年人，测试在这些人家里进行，也发现了“频率-效力效应”，而且其程度与实验室中美国大学生被试所得的几乎相同。

在第 10 章中，我们将讨论许多带有概率性质的决策原则，这些决策原则最早都产生于实验室，但都经过了现场式的检验。例如，研究者检验了理疗师、股票经纪人、陪审员、经济学家及赌徒在各自所属情境下，是以何种方式做出概率推理的（Belsky & Gilovich，1999；Gilovich，Griffin，& Kahneman，2002；Hilton，2003；Kahneman & Tversky，2000）。行为决定理论的原则业已用于许多应用性的领域，例如，决定丹佛市警局使用哪种型号的子弹最为理想，以及决定是否在亚利桑那州中部建立大坝（Hammond，Harvey，& Hastie，1992）。

实验与非实验结果的聚合性也成为教育心理学领域的突出特点。例如，针对不同课程安排所做的实验研究和现场研究都表明，早期语音教学有助于阅读技巧的习得（Ehri，Nunes，Stahl，& Willows，2001；Pressley，2005；Snowling & Hulme，2005；Stanovich，2000；Vellutino et al.，2004）。

总的来说，当前的心理学研究采用多种类型的实验技术和情境。尽管对于很多问题的研究有时候过于集中在使用某些特定的技术，但在心理学中，研究方法的多样性比过去几年丰富多了。

向更有效的研究方法迈进

对于某个特定问题的研究，通常是从相对较弱的方法过渡到可以做出较强结论的方法。例如，研究者对某个特定假设的兴趣，常常源于某个异常感兴趣的特殊个案。正如我们在第 4 章中讨论的，这就是个案研究的真正作用：为更有效力的进一步研究提供一些假设，同时激发科学家们用更为严格的方法去研究这些假设。个案研究之后，研究者多采用相关研究来确认变量之间是否存在真正的关联，而不仅是存在于几个个案中的巧合现象。如果相关研究证实了变量之间的关联，研究者就开始尝试采用实验法来对相关变量进行操纵，借以找到变量之间可能存在的因果关系。这个递进的顺序就是：从个案研究到相关研究，再到操纵变量。尽管并非每个研究领域都遵循这个渐进式途径（有时不同类型的研究同时进行），但这一向更有效方法迈进的进程的确是普遍发生的。

在讨论“向更有效的研究方法迈进”之前，我们必须纠正读者的一个错误概念，这个错误概念源于第 5 章的讨论，那就是“相关研究在科学中没有什么用处”。的确，当一个因果关系的假说需要验证时，操纵变量的研究方法更受青睐。然而，这并不意味着相关研究对于知识的获得毫无帮助。首先，许多科学假设是以相关或者不相关的形式来表述的，因此这类研究是在直接验证这些假设。第二，尽管相关并不意味着因果关系，但因果关系一定包含相关。也就是说，如果一个相关研究不能肯定地证实因果关系的假设，那它可以起到排除这一因果假设的作用。第三，相关研究或许比它们看上去更有用，因为最近新发展的复杂相关设计可以让研究者做出有限的因果推论。我们在第 5 章讨论了偏相关这种复杂的相关技术，这一技术有可能检验出变量间的关联是否能够被第三变量所解释。

然而，最重要的原因可能在于，有时出于道德或伦理的考虑，我们无法对一些变量进行操纵（例如，营养不良或肢体残障）。而另外一些变量，诸如出生顺序、性别、年龄等，则因其无法被操纵而具有天然的相关性，涉及它们的科学知识也因此必须建立在相关证据基础上。当然，这一情况并不是心理学领域所独有。天文学家们显然无法操纵所有影响其研究对象的变量，然而他们依然能够做出结论。

在健康心理学中，有一个研究方法演进的例子，它涉及 A 型行为模式和心脏病之间的关系（Curtis & O'Keefe，2002；Matthews，2005；Smith，2003；Suls & Bunde，2005）。最初，A 型行为模式这一概念源于两位心脏病专家的观察，这二位医生从他们一些病人的行为中发现了一种稳定的模式，这种行为模式包括时间紧迫感、飘忽不定的敌意，以及对成就的极度渴求。于是，一些医生通过对少数个案的观察，提出了“A 型人格”这一想法。这些个案研究提出了这个概念，但并不能作为有力证据来证明这种特定的行为模式是导致心脏病的原因之一。要证明这一点，需要的不仅是少数几个个案研究，它还需要由心脏病专家、生物化学家和心理学家团队数十年的努力。

很快，这个研究从永远也不可能证实假设的纯粹个案研究，转向了更有效力的研究方法。研究者发展和检验了 A 型行为模式的操作性定义。大范围的流行病学研究证实了 A 型行为和心脏病之间的相关性。然后这种相关研究工作就变得很复杂了。研究者使用复杂的相关技术来搜寻潜在的第三变量。由于行为模式与其他传统心脏病风险因素中的一种（例如吸烟、肥胖和血液中胆固醇水平）存在相关，因此 A 型行为和心脏病之间有可能存在虚假相关。当其他的变量在统计上被排除后，A 型行为模式和心脏病之间仍然具有关联。

最后，研究者采用了实验研究对变量进行操纵，以期证实二者间是否具有因果关系。一些研究试图去验证是否某些生理机制影响了两者之间的关系，并以动物作为被试——某些人所谓的“不是真实的生活”的研究方法。另外一些研究则以犯过心脏病的人为被试。这些被试被随机分配到两个组中的一组。一个组接受咨询，帮助他们避免传统的风险行

为，例如吸烟或者吃高脂肪食物；另一组在接受同样的咨询的同时，还接受了一个以减少他们的 A 型行为为目的的训练项目。三年之后，在接受 A 型行为辅导的病人中，心脏病复发的情况要明显少很多。

简而言之，证据汇聚起来支持了“A 型行为模式是导致心脏病的重要原因”这一假设。对这个问题的研究提供了一个很好的范例，从中我们能清楚看到，研究是怎样从一个感兴趣的个案研究走向相关技术，最后到可以操纵变量的实验研究的。

我们能从这个例子中得到的最后一点经验就是，科学概念总是在不断地演进。这个论点是在第 3 章讨论操作性定义时首次提出的。最近的研究似乎表明，将 A 型行为与心脏病之间的关系说成是整体性的显得过于简单化了。原因在于，只有该概念中的特定成分（特别是对抗性敌意）才与心脏病有关联（Curtis & O’Keefe，2002；Matthews，2005；Suls & Bunde，2005）。因此，这是个很好的例证，从中可以看出，随着科学的进步，它是如何不断地揭示特定的关联，以及理论概念是如何被细化的。

不要对矛盾数据感到绝望

聚合性原则的最后一个启示是，当一个问题的最初的研究结果看上去有些矛盾时，我们不应当对此感到绝望。在科学中，证据融合的过程就像投影仪慢慢将一张未知的幻灯片的焦点调清晰。起初，屏幕上的模糊影像可能代表任何东西。接着，随着一点点地调整焦距，虽然这个图像仍不能被清楚地识别出来，但许多其他的可能假设也许会被排除。最后，当焦距调准，就可以非常有信心地做出最终的判断。证据融合过程就好比一个调焦过程。幻灯片的模糊影像就如同互相矛盾的数据，或者是那些支持多重假设的证据。

因此，研究早期所获得的矛盾数据不应该让我们对发现真相感到绝望。类似的情况不光发生在心理学领域，同样也发生在一些相对成熟的

科学中。的确，公众经常意识不到科学中经常会得到一些矛盾的数据。这些矛盾只不过是因为我们对问题理解得还不够充分，这些矛盾还可能仅仅是偶然事件（我们将会在第 11 章中对此展开讨论），或者源于不同实验在方法上的细微差异。

在达成共识之前，其他许多科学也都经历了令人困扰的不确定时期（Ioannidis，2004；Simonton，2004）。格兰德威尔（Gladwell，1996）讲述了近来关于脑创伤患者紧急救治认识的演进过程。一名纽约患者非常幸运地得到了世界顶级专家之一简姆·加哲医生（Dr. Jam Ghajar）的治疗。这位医生始终致力于改变该领域中一个临床上的错误看法。格兰德威尔说，若干年前，当加哲和其他五位研究者在对一些创伤治疗中心进行调研时发现，尽管类固醇已经被反复地证明无助于减少颅内压力（而且会带来潜在危害），然而仍有 75% 的昏迷状态的病人是用类固醇来治疗的。他写道：

> 部分问题在于，在神经外科领域里，很难在治疗过程或者治疗上形成可信的、科学的结论。再者，大脑的复杂性和神秘性导致了一种提倡直觉的文化，并让每一个神经外科医生坚信自己的经验与其他任何人的经验一样有效（p. 39）。

当谈到几年前他的同行的观点时，加哲说，并不是说神经外科医生太过懒惰，而是这儿的信息太过庞杂，让人感到困惑（p. 39）。

简而言之，和心理学的众多领域一样，该领域也充斥着许多尚未聚焦的研究，并且未能以一种有利于找到聚合性的方式加以概念化。因此在 1994 年，加哲和他的几个同事参加了一系列学术会议，在这些会议上，他们试图对所有的证据进行综合，以期发现某种聚合性。这些会议是由大脑创伤基金会发起的，研究者们审阅了涉及大脑创伤处理的 14 个领域、超过 4000 份的科学论文。大脑创伤基金会的执行主席描述了神经外科专家是怎样工作的：“他们所做的工作是对科学文献的证据进行论证，一旦有人说‘我以往的经验就是如此’，所有人都会说，‘哦，不，那不算，我们要看到证据’”（Gladwell，1996，p. 40）。最后的结果被证

明是富有成效的：

> 这个团体确实在文献中发现了聚合性，并出版了一本专著来展示所有的科学证据以及脑创伤护理各阶段的最新治疗手段。这本指导手册是神经外科界首次成功建立的一套标准治疗规范，如果美国大多数创伤中心都能严格按照这些规范去做，每年就可以挽救一万名以上的病人。（Gladwell，1996，p.40）

这个指导方针已经用于挽救生命，但是有趣的是，当加哲自己开始讨论这段医学史的时候，他强调了证据聚合原则的价值和科学知识的公共性（来自本书第1章的一个原则）。“人们想把它归功于具体的某个人，我猜这是人类的天性，他们想要说，‘这是加哲创立的规范，他是一个伟大的医生’，但事实并非如此。这些标准是建立在当前最好的科学证据之上的，而且每一个人都可以使用。”（Gladwell，1996，p. 40）

从这个例子可以明显看出，并不仅是心理学的研究领域中遍布各种发现，在其他科学中也不乏这种由于数据模式的模糊性而难以达成一致结论的例子。在一篇名为“图片问题”的文章中，格兰德威尔（Gladwell，2004）讨论了人们为何很难理解医生对于乳腺X光片的作用还存在着分歧。这是因为乳腺X光透视在大多数人看来是如此地“精确有力”，以至于他们认为仅凭它就能做出确诊。其实这些人不理解，医生的诊断虽必不可少，但乳腺X光片评估和疾病预测从本质上来说是具有概率性的。格兰德威尔说，“图片保证确定性，但它不能兑现这种承诺。经过40年的研究之后，对于女性在50岁至69岁期间接受乳腺X光透视的益处，仍然存在着不小的分歧。进一步的争议则在于，是否有足够的证据能够证明，50岁以下和70岁以上的女性定期需要接受乳腺X光透视检查”（p. 81）。然而格兰德威尔继续谈到，和心理学领域一样，在医学领域里，知识即使不确定也依然有用：“答案是乳腺X光透视不需要完全准确无误才能拯救生命……它没有我们想的那么好。但总归比没有它要强”（p. 81）。

在心理学和其他科学里，将来自不同研究的证据整合起来形成一

个结论，已经能够通过一种更为正式的方法来实现，这就是一种叫做元分析的统计技术（Cooper & Hedges，1994；Hunter & Schmidt，1990；Rosenthal，1995）。在医学语境中，元分析就是：

> 将许多临床实验的数据汇总起来，形成一个足够大的数据库，以消除困扰单个实验的统计不确定性。元分析的最大优点，就是能从结果遍布的多个实验之中产生出一个明确的结果。（Plotkin，1996，p. 70）

使用元分析来确定心理学结论的有效性，和医学的情形是类似的。两个实验组的比较得出的效应，可以纳入一个常规的统计矩阵中，这个矩阵能进行研究之间的比较。接着，这些结果以一种标准化的方式加以统计整合（Cooper & Hedges，1994；Hunter & Schmidt，1990）。如果整合过程达到了一定的统计学标准，就能形成一个关于这些效应的结论。当然，在某些情况下，有可能无法确定地得出一个结论，这时元分析的结果就是非结论性的。

越来越多的评论者开始呼吁，应更加重视元分析，并将之视为一种方法，来消除科学领域内相互对立研究之间的不断争议。这种方法有助于终止这种“公说公有理，婆说婆有理”的争论。对元分析的强调也揭示了一种观点：专业杂志上常见的观点对立可能只是表面现象，实际上我们拥有更多可靠和有用的发现。

国家阅读评审小组（NRP，2000；Ehri，Nunes，Stahl，& Willows，2001）对一些关于阅读教育的研究所做的元分析就证明了这一点。例如，他们得出结论，对 38 个不同的研究结果的元分析“有力地支持了这一观点，即相比其他课程提供的非系统或非语音教学，系统的语音教学在孩子的成长中发挥了更大的作用”（p. 2-84）。在报告的另一部分，NPR 报告说，对于 52 个语音意识训练研究的元分析说明，“教孩子掌握在语言中运用声音，能帮助他们学会阅读，在不同的教学、测验及参与者的个性条件下，其效应量都远远大于随机水平，并且，虽然这些效应有大有小，但大部分都处于中等水平”（p.2-5）。

美国心理学会的一支工作团队在心理学期刊上所做的关于统计方法的一番阐述，为本节内容提供了一个恰当的总结（Wilkinson，1999）。这个工作团队说："研究者不应仅针对单个研究的结果做出解释，就好像其他文献所报告的结果与之毫无关系似的"（p.602）。不同研究结果之间达成聚合效应，才有利于推动科学进步。一个研究的结果也只有通过针对特定问题的诸多研究获得聚合性解释，才是有意义的。

小 结

在这一章中我们看到，为何“跃进”模式对于心理学来说是一种糟糕的模式，以及为什么“渐进整合”模式提供了一个更好的框架，凭借这个框架，我们就能够理解心理学中的结论是如何形成的。聚合性证据原则描述了心理学上研究结果是如何被整合的：没有一个实验是可以一锤定音的，但是每一个实验至少都能帮助我们排除一些可能的解释，并让我们在接近真理的道路上向前迈进。多种不同方法的使用让心理学家更为确信，他们的研究结果是建立在稳固的实证基础上的。最后，当概念上的变化发生时，它必须遵循关联性原则：新的理论不仅要能解释新的科学数据，还必须能解释已有的数据。

9

打破“神奇子弹”的神话
——多重原因的问题

在第 8 章里我们关注了聚合性操作的重要性，以及寻求一种能够在变量间建立单一联系的更有效的研究方法。在这一章中，我们将不只看两个变量间的单一联系，还将重点关注另一个重要的观点，那就是人的行为是由多重原因共同决定的。

任何一个特定行为都不是由某个单独的变量引起，而是由许多不同的变量共同决定的。认定变量 A 和行为 B 之间存在显著的因果关系，并不意味着变量 A 就是引起行为 B 的惟一因素。例如，有研究者发现，收看电视的时间和学业成绩之间存在相关，但不会就此认为收看电视时间是影响学业成绩的惟一因素。道理很简单，学业成绩在一定程度上还受到大量其他变量的影响（例如，家庭环境、学校教育的质量等等）。实际上，相对于这些变量，看电视只是影响学业成绩的一个次要因素而已。同样地，收看大量的电视暴力也不是使儿童表现出攻击行为的惟一原因，它只是众多影响因素中的一个。

但人们常常忘记行为是由多重原因决定的，他们似乎要去寻找那颗所谓的“神奇子弹”——即他们感兴趣的、造成行为的惟一原因。心理学家希尔多·瓦茨（Theodore Wachs，2000）以人们试图解释 1998 至

1999年间发生在美国的校园枪击案的方式作为例子，指出，人们认为涉及的原因包括枪支容易获得、父母对孩子较低的关注、互联网、影视暴力、同伴影响和精神疾病。瓦茨认为，“很少有人觉得校园枪击案激增是上述原因共同作用的结果，任何解决方案都不应只针对某一个潜在的原因”（p. x）。

和本书中谈到的许多其他原则一样，具备原因多样性的观念非常重要。一方面，它提醒我们不要过于依赖单一的原因解释。因为这个世界盘根错节，影响行为的因素也多样而复杂。虽然我们可以证明某一变量引起了某一行为，但并不代表已经发现了影响该行为的惟一原因，甚至是最重要的原因。为了对某种特定行为做出全面的解释，研究者必须探讨各种不同的变量对它的影响，并把这些研究结果整合起来，才能完整地描绘出所有与该行为有关的因果关系。

另一方面，虽然说某个变量只是影响特定行为的众多因素之一，并且只能解释这一行为的一小部分，但并不是说这个变量就是无足轻重的。首先，这一关系可能具有深远的理论意义。其次，这一关系可能具有应用价值，尤其当这个影响变量是可以进行人为控制的时候，如前面提到的电视暴力的例子。如果控制了这一个变量，能够使每年的暴力事件降低1%，那我想没有人会认为它是无关紧要的。总之，如果问题行为至关重要，那么懂得如何去控制其中一个哪怕非常小的原因也具有非凡的价值。

罗森塔尔（Rosenthal，1990）举过一个治疗心脏病的例子，在一个实验中，某种治疗方案能将患者存活率提高不到1个百分点；然而，即使这样，这个结果也被认为是意义太过重大，以至于基于伦理考虑，实验者不得不提早终止研究：既然实验治疗结果这么有效，对那些被随机分配在控制组的病人，让他们仍然使用安慰剂显然是违背伦理的。同样，任何能够将机动车死亡率降低1%的因素都至关重要——每年都能挽救450条生命。将凶杀案案发率降低1%，则每年能挽救超过170条生命。总之，一个结果是由多重变量决定的这一事实，并没有降低任何一个与结果存在因果相关的变量的重要性——即使这一变量仅能

让结果产生很小的变化。

交互作用

原因多样化的观点引出了另一个重要概念，那就是交互作用。这个概念在许多方法论的书上都有详细的介绍，因此这里不再赘述，只是稍提一下：当影响行为的一个因素和另一个因素共同起作用时，它们会对该行为产生与各自单独起作用时截然不同的影响。这就是我们常说的交互作用：一个自变量的影响效果依赖于另外一个自变量的不同水平。赛蒙斯等（Simmons，Burgeson，Carlton-Ford，& Blyth，1987）指导的一个实验提供了这样的一个例子。这些研究者考察了一组青少年的学业平均成绩，想看看一些生活事件（如转学、青春期发育、早恋行为、搬家和家庭破裂等）是否会对学业产生影响。他们发现上述生活事件加在一起，是导致学业不良的关键因素。

另一个例子是迈克尔·努特（Michael Rutter，1979）对儿童精神疾病相关因素方面的研究进行的综述，他提出：

> 第一个引人注目的发现是，在实验中，那些被单独分离出来的慢性压力并未增加精神疾病的风险……这些风险因素单独作用时，没有一项与儿童的精神疾病存在关联；这些儿童患精神疾病的风险也不会比没有家庭压力的儿童高。然而，当任何两种不同来源的压力同时作用时，患病的风险就超过原来的 4 倍。若是 3 种或 4 种压力来源同时作用，那么患病的风险更是增大了好几倍。很明显，这些慢性压力的共同作用远远超过其各自效果的累加，因为几种并发压力之间存在交互作用，才令其总体效应远远大于单个压力效应之和。（Rutter，1979，p. 295）

当诸如努特所描述的交互作用发生时，要理解其发生的逻辑，可以先想象一个风险量表，得分 80~110 代表低风险，110~125 代表中等风险，

125–150 则代表高风险。假设我们发现儿童在无压力情况下的平均风险得分为 82，在压力因素 A 作用下的平均风险得分为 84，而在压力因素 B 作用下的平均风险得分为 86。当研究因素 A 和因素 B 两者对儿童的共同影响时，如果发现风险指数达到了 126，也就是说，联合的风险指数远远超过了独立研究单一因素时所预测的结果，就说明了因素 A 和 B 之间存在着交互作用。

发展心理学中也有许多类似努特所描述的例子。邦尼・布瑞特米亚和克雷格・拉米研究了两组婴儿，一组是非最佳围产期的婴儿，另一组是正常婴儿（Breitmeyer & Ramey，1986）。在这两组婴儿出生后，再把他们随机分配成两组——实验组及控制组，然后对实验组实施一个特别的育婴方案，该方案是为了防止出现轻微智力迟缓而设计的。控制组的婴儿则没有得到任何特殊的照料。当这些孩子长到 4 岁的时候，对他们的认知发展能力进行测试，发现在特别育婴方案下，非最佳围产期出生的儿童与正常儿童在认知能力上没有显著差异。但是，没有得到特殊照料的控制组中，那些非最佳围产期儿童的表现低于正常儿童的认知发展水平。该研究中，生理和环境因素的交互作用说明，一个复杂的行为结果（认知发展）是由多种因素决定的。当非最佳围产期出生的儿童得不到适当的照顾时，就会出现负面的认知发展结果。研究者们总结道：“这个研究结果支持了这样一个理论架构，即对于那些在社会经济条件较低的家庭中成长的儿童而言，先天的生理缺陷和后天不良的环境因素会成为他们发展中的累积性危害因素”（p. 1151）。

同样地，另一项有关验证抑郁症的“体质－压力理论”的研究（Metalsky & Joiner，1992）表明，消极生活事件与三种脆弱的心理因素共同作用，导致罹患抑郁症的可能性达到最大。这三种因素是：归因风格（倾向于把消极事件归因为一些稳定的、整体的因素）、对自我作消极推断，以及对任何行为后果进行消极推断的总体倾向（见 Alloy，Abramson，& Francis，1999）。

很多消极的行为和认知后果都伴随着相似的逻辑。例如，儿童攻击行为是由遗传基因和不良社会环境相互作用所造成的（Pennington &

Ozonoff，1996）。同样地，佩迪特等（Pettit et al.，1999）发现，青少年在成长的早期如果花费大量时间与同伴交往、同时在家里也很少受到父母监管的话，会更容易出现外化行为问题。

积极的结果也可用多种因素及其之间的交互作用来解释。耐特等（Knight et al.，1994）在研究 6~9 岁儿童的亲社会行为时，检验了与儿童助人行为倾向（如，捐款给有需要的儿童）相关的心理因素。他们发现一些变量——如同情心、情感推理和关于金钱的知识等——单独作用时，它们和亲社会行为之间的相关很低。但是，当这些变量联合作用时，能够很好地预测亲社会行为。例如，具有较强的同情心、较强的情感推理并对金钱有所认识的儿童，捐款的数目是在这些变量上表现较低的儿童的 4 倍。

因此，原因多样化的概念可能比你最初设想的要复杂得多。不仅需要追踪并测量影响问题行为的种种可能因素，还必须考察这些变量是如何共同作用的。

单一原因解释的诱惑

复杂事件是由多重原因所决定的，这个基本的理念似乎很容易理解。实际上，当问题没有太大争议时，这个观点确实很容易掌握和运用；但是，当预设偏见——这个科学工作者的老敌人（参见第 3 章）开始抬头时，人们就会倾向于忘记原因多样性这一原则。我们无数次听到，人们对一些容易引发情绪的话题——如犯罪的原因、财富的分配、对妇女和未成年人的歧视、贫困的原因、死刑的作用以及纳税标准的争论等，都在以这样一种方式进行，那就是让人觉得这些问题是简单的、单维的，而且导致结果的原因只有一个。这些例子进一步反映了尼斯贝特和罗斯（Nisbett & Ross，1980）的说法：“虽然人们有时承认原因多样性，但是，人们在行动上却更多与单一原因的信念相一致。在某种意义上说，人们的行为像是将各种原因视为“水压式”的，或者各种可能原因之间

像在‘零和’游戏中那样相互竞争”（p.128）。

“零和”游戏—— 一个人的收益是另一个人的损失——常常反映了我们如何讨论那些容易引发情绪的话题。在情绪的影响下，人们通常会忘掉原因多样性这一原则。想想两个敌对的政党是如何讨论社会犯罪问题的。自由主义者会认为那些社会经济地位低下的人之所以会犯罪，是因为他们本身就是恶劣社会环境（如失业、恶劣的住房条件、缺乏教育和对未来丧失希望等）的受害者。而比较保守的人会争辩说，也有许多穷人并没有犯罪，所以社会经济条件并不是主要原因。与之相反，他们认为个人的价值观和人格特征才是决定犯罪行为的真正原因。双方似乎都没有认识到个体因素和环境因素共同导致了犯罪行为。

政治评论家理查德·科恩（Richard Cohen）也曾在文章中谈到我们通常是如何根据预设偏见来将“单一原因”的解释调转 180 度的。他引用了爱荷华州一名 63 岁农场主的例子，该州经历了严重的农业衰退，该农场主因此背负上了沉重的债务，并失去了他的农场。在感到还债无门的绝望中，他开枪打死了他借钱的那家银行的经理，随后枪杀了他的妻子，之后自杀。邻居和传统媒体的看法都认为，这个农场主是因为身陷巨大的财务困难才会“崩溃”的。媒体的报道都对他报以极大的同情。科恩写道，这位农场主被描绘成一位“勤劳的企业主，他顽强地与大自然、银行及芝加哥农产品商人搏斗。他诚实而节俭，具备大多数美国人的优点——个体经营、自食其力，而且敬畏上帝”（1985，p. 11）。

但是科恩疑惑的是，如果这个人真的只是因为财务困境就去杀人和自杀的话，那么我们是否可以用同样的（单一原因）理论去解释那些发生在贫民窟的杀人事件？“如果对于受困的农场主而言，这是其崩溃的原因的话，那么为什么发生在贫民窟的情况就不同了呢？为什么有人提出贫穷、缺少机会、三流学校和野蛮行径是造成犯罪的原因时，会受到如此多的责难？”（p. 11）。当然，科恩指出了在作单一性原因解释时产生的另一种谬误：我们利用单一解释来巩固已有的偏见。科恩认为，如果从一开始就认识到上述例子中农场及贫民窟的杀人事件都可能是由多样因素决定的，或许就能避免这一谬误。不论是爱荷华州的农场主还是

贫民窟的杀人行为，都同时受到了个人的心理 – 生理特性和环境压力的共同影响。没有一种单一的原因可以单独解释犯罪事件。犯罪行为是受许多因素影响的，部分是环境因素，部分是个体因素。

对复杂的经济形势的讨论提供了这样一个例子。一个曾引起了长达几十年争论的、极具社会意义的重要话题是：为什么美国的贫富差距不断扩大（Beatty，1996；Bronfenbrenner et al.，1996；Frank，1999；Karger，2005；Madrick，2006）。正如本书在第 6 章中所提到的“聪明汉斯”的例子一样，这一事实不存在争议，有争议的是对这一事实的解释。1979 年以来，美国男性工人的实际收入（即排除通货膨胀的影响之后）下降幅度超过了 10%（Cassidy，1995；Mishel，1995）。当然，也有个别群体状况不错。从 1977 年到 1990 年的十余年间，人口总数里收入最高的 1% 的人，收入增长了 74%（这个数值也是在排除了通货膨胀之后的数据，Slemrod & Bakija，1996）；与此同时，美国中产阶级的收入只增长了 3%；而占人口总数 20% 的处于社会底层的人，他们的收入却下降了 13%。1977 年，社会上最富的 20% 的人所赚的钱是那些最穷的 20% 的人所挣得的 4 倍。而到了 1991 年，这一数字是 7 倍（Frank & Cook，1995）。

财富从公民的一个阶层大规模地转移到另一个阶层手中，这一现象引发了一场极富争议的、有关其原因及影响的政治辩论。这场争辩最引人注目之处就是，这些争论者都只关注单一的原因。争辩的每一方都只以某一个原因为立论基础，然后千方百计地攻击所有支持其他原因的观点。事实上，计量经济学研究（Beatty，1996；Cassidy，1995；Frank & Cook，1995；Mishel，1995）已经聚焦了四个变量（还有人提出了超过四个的变量，但这四个是得到最广泛关注及研究的）。其中一个因素是科技。例如，论据之一是，计算机提高了其使用者的生产效率，导致了其收入的增加。与之相反，计算机还替代了很多没有特殊技能的工人（如邮件分类员、银行出纳员等等）的工作，从而降低了他们的工资收入。争论的第二个因素是新移民不断涌入美国，而这些人多是非熟练工，他们造成了非熟练劳动力供大于求，使得已经很低的工资水平继续下滑。

第三个原因是全球化，它进一步加剧了收入不均，因为公司可以通过业务外包，在一些工资水平较低的国家雇佣一些非熟练工和半熟练工（正变为熟练工），而这更加重了本国非熟练劳动力的过剩。第四个原因是工会和大企业在影响力上的此消彼长。论据是，在 20 世纪 90 年代，劳工的罢工事件减少了，而资方的罢工事件（即企业放弃了某一区域为其生产基地，因为在其他区域生产可以取得更高的投资回报率）却增加了，从而在提升资本价值的同时降低了劳动力价值。

经济学研究这四个变量的时候到底发现了什么呢？你已经猜到了。所有这四个因素共同作用造成了不断加重的社会不平等。这个例子也证明了先前所提到的交互作用的概念。卡西迪（Cassidy，1995）在文章中指出，“某些因素可能有交互作用并相互强化。随着全球化竞争日益激烈，企业管理者削弱工会，将资本投入计算机技术。同样地，公司迁址的威胁和外来廉价劳动力的增多都会导致工会力量的进一步削弱”（p.122）。

和经济学的问题一样，心理学所研究的几乎所有复杂问题也都是由多重原因决定的。以学习障碍为例，这个问题已经被教育心理学家、认知心理学家和发展心理学家广泛地研究过。结果发现，脑部的病变与学习障碍有关（Galaburda，1994；Hynd，Clinton，& Hiemenz，1999）。还有研究发现，学习障碍具有遗传方面的原因（Cardon et al.，1994；Olson，1999）。这两个研究结果看起来好像可以让我们做出一个结论：学习障碍是纯粹的生理 - 脑的问题，但这样的结论是错误的，因为进一步的研究发现，造成学习障碍的部分原因是在早期学校教育中缺乏某些指导性的经验（Pressley，2002），以及贫穷的家庭环境（Senechal & LeFevre，2002；Snow，Burns，& Griffin，1998）。学习障碍因此不是由单一原因所引起的；相反，它是生理与环境因素交互作用的结果。

小　结

本章内容虽然简单，但却非常重要。考察行为的原因时，要依照多样性的原则来思考。不要陷入误区，认为某一特定行为只是由某一特殊原因造成。大部分复杂的行为都是由多重原因所决定的。各种各样的因素共同起作用才导致了某种行为出现。有时多个因素联合在一起时会产生交互作用。也就是说，变量共同作用时的整体效应，会和其单独作用时获得的效应完全不同。

10

人类认知的阿喀琉斯之踵*

——概率推理

问：男人比女人高，对吗？

答："对。"

问：所有男人都比所有女人高，对吗？

答："错。"

完全正确。信不信由你，在这一章里，我们还将花一些篇幅来讨论你已经知道的一些问题，这可以从你刚才回答上面两个问题时看出，你已经知晓了一些答案，但是，先别因此就跳过这一章。因为接下来在我们对一些看似非常简单的原则所作的解释之中，会有惊喜等着你。

你为第一个问题给出了肯定的答案，这是因为你没有把"男人比女人高"这句话理解成第二个句子所说的"所有的男人都比所有的女人高"。你把第一句问话正确地理解为"男人有比女人高的趋势"的意思，因为每一个人都知道，不是所有的男人都比所有的女人高。你理解到那句问话反映了一个概率趋势，而不是一个在任何情境中都适用的事实。我们

* 意指致命的缺点或要害——译者注。

所说的概率趋势是指有较大的可能性，但并非在所有情况下都必然如此。也就是说，性别和身高的关系要用可能性和概率的词汇来描述，而不是用必然性的字眼。在自然界中很多关系的本质也是概率性的，例如：接近赤道的地区比较热；每家的孩子数目不超过 8 个；地球上大部分地区昆虫的数量比人类多。这些都是统计学可证明的趋势，但是它们当中的每一句话都不是绝对的，仍然可能会有例外。因为它们是概率的趋势和规律，而不是在所有情况下都成立的关系。

事实上，心理科学所揭示的所有事实和关系都是用概率来表述的。这一点也并非心理学所独有。在其他学科里，很多定律和关系也是用概率而非必然性来表述的。例如，人口遗传学的所有子学科都基于概率关系；物理学家告诉我们，原子中电子负荷的分布也是通过概率函数来描述的。

确实，心理学所揭示的大部分概率趋势都比较弱。在心理学中，各种行为关系都是以概率形式加以描述的，然而这一事实并没有使得它与其他科学之间产生天壤之别。正如雅各布·布朗诺斯基（Jacob Bronowski）所言（1978a），许多人还是无法接受这样一个事实，那就是随着科学不断地开拓出新的研究领域，越来越多的科学定律都将采用概率形式加以描述：

> 如果我说，在经过了风和日丽的一周之后，周日总要下雨，这会被认为是一个规律。但是，如果我说，在经过了风和日丽的一周之后，星期天下雨的可能性比不下雨的可能性要大，这就是一个不太令人满意的说法，并且人们会理所当然地认为，我没有真正发现一个潜在规律，这种对规律的寻求符合我们的一种习惯，即想让科学说出决定性的“是”或“否”。甚至如果我说，一周中，经过前 6 天好天气之后，10 个周日里有 7 个会下雨，你可能会把它当作一个统计数字来接受，但是它还是不能让你满意，因为它不是一个定律。它看上去多少缺乏如规律那样的力度。然而这纯粹是一种偏见。我所解释的概率的概念并不难，但它新鲜而陌生。我们不习惯去面

对它…… 我们似乎生活在“有时”和“或许”的世界里，但希望与“始终”和“确定”为伴…… 我也相信这一困难不是一种习惯。一旦我们愿意或者不得不接受这样的理念，我们就会尽快适应它。其实我们不得不这样做。（pp. 81–82，94–95）

在这一章里，我们想尽可能地让你在这个“有时和或许的世界”里感到更舒服一些，因为，一个人若想要理解心理学，就必须对“概率推理”这一本章的主题安之若素。

“某某人”统计学

大部分公众都能意识到，医学的许多结论采用的都是概率趋势而非绝对确定性的表述。吸烟会导致肺癌并诱发其他健康问题。相关的医学证据汗牛充栋。但每个吸烟者都会得肺癌吗？所有戒烟者都解除了患肺癌的风险吗？大多数人都不会认为这些推论能够成立。吸烟很大程度上增加了患肺癌的概率，但并非绝对。医学能够以很大的把握告诉我们，吸烟群体中的人比与之相似的非吸烟群体中的人更容易死于肺癌，但不能告诉我们是哪一些人会死，这种关系就是概率；它并不适用于所有个案。我们都知道这一点——真的知道吗？我们经常看到下面这样的场景：一个不吸烟的人引用吸烟导致肺癌的统计数据，试图说服一个瘾君子戒烟，所得到的结果仅仅是对方的反唇相讥：“嘿，走远点儿！你看那个铺子里的老乔，他从 16 岁开始，每天要吸三包骆驼烟！现在他已经 81 岁了，看上去还很结实！”人们对此可能做出的推断显而易见：就是这一个特例已经推翻了吸烟和肺癌之间的关系。

令人吃惊和沮丧的是，这种反驳手段屡试不爽。通常情况是，每当一个个案被用来证明概率趋势无效时，很多人都常常点头表示赞同，这反映出他们没有正确理解统计规律的本质。如果人们认为一个特例就可以让一个规律失效，他们一定认为这个规律应该在任何情况下都适用。

简言之，他们错误理解了概率定律的性质。既使是最强的趋势也会有少数的“特例”与之相悖。就拿吸烟的例子来说，活到 85 岁的人中只有 5% 是吸烟者（University of California，Berkeley，1991）。或者从另一角度来看，活到 85 岁的人中有 95% 属于从不吸烟者，或在一段时期内吸烟但最终戒断者。连续从未间断地吸烟会显著地缩短寿命（Univer-sity of California，Berkeley，1991），然而也有少数吸烟者活到了 85 岁。

借用心理学家尼斯贝特和罗斯（Nisbett & Ross，1980）的术语，我们把类似“老乔”的故事称作“某某人”统计学的运用：由于某些人知道一个“某某人”与某个成熟的统计学趋势相左，这个趋势就会被人怀疑。例如，我们经常听到类似的话——“你是说服务业的就业机会正在扩大而重工业中则在缩小？这不对，我就知道‘某某人’上周四在一个钢铁厂找到了一份工作”；“你说与 30 年前相比，家里的孩子少了？少胡扯！隔壁的年轻夫妇已经有了 3 个小孩，但他们还不到 30 岁”；“你说通常孩子都会倾向于信仰他们父母所信仰的宗教？但据我所知，我的一个同事的孩子就在前几天改信了另一门宗教。”

当我们面对和过去持有的观念相矛盾、同时又是强有力的证据时，无所不在的“某某人”总是会立刻跳出来否定这些统计规律。因此，我们可以说，实际上人们知道的不少，他们只不过顺手把“某某人”当成一种工具，把与他们观念相悖的事实给否决掉而已。然而，研究人类决策和推理的心理学家们的研究结果表明，人们之所以使用“某某人”，不只是由于它是一个有用的辩论手段。相反，这一错误的争论模式之所以被应用得如此频繁，主要在于人们不知道如何处理概率信息。决策心理学的最新研究发现，概率推理可能正是人类认知的阿喀琉斯之踵。

概率推理以及对心理学的误解

科学、技术和人事等许多领域都涉及概率思维。所以，我们也没有什么特别的理由认为这种思维对理解心理学比其他学科更重要。然而，

由于人们在运用概率信息方面存在问题，导致心理学的研究结果常常被误解。我们都理解“男人比女人高”是一个概率趋势的陈述，所以并不会因为有一个特例（某个男人比某个女人矮）就认为这一陈述是错的。很多人也能以同样的方式来理解“吸烟可以导致肺癌”的陈述，尽管对于那些不愿相信吸烟会导致其丧命的瘾君子们来说，“老乔”可能还是有说服力的。然而，与之相似的有关行为趋势的概率表述却引发了广泛的猜忌，而且常常是“某某人”刚一露头，这种概率表述便被人们抛弃了。很多心理学教师在讨论某些行为之间关系的证据时，都往往得到同样的反应。例如，教师可以呈现如下的事实：儿童的学业成绩和家庭的社会经济地位及父母的教育水平相关。但这个事实常常会遭到至少一个学生的反对，他会说，他有一个朋友是国家优秀奖学金获得者，但是他的父亲只是中学毕业。甚至那些理解吸烟－肺癌例子的人，对这一问题的态度也变得摇摆不定了。

人们从没想到过要用“某某人”的论据来反驳医学和物理上的发现，却习惯于用之驳斥心理学的研究结果。大多数人能理解医学科学提出的治疗、理论及事实是概率性的。例如，他们理解一种药对一组病人来说，并不是对他们各个都有疗效，而且医学也经常不能事先告诉我们，该药会对哪些病人有疗效。通常可以说，100 个病人接受某治疗方案，100 个病人不接受任何治疗，在一段时间之后，接受治疗的这 100 个病人总体来说会比不接受治疗的 100 个病人的病情好转一些。没有人因为这个并非在所有情况下都适用的概率表述就怀疑这一治疗的价值。许多心理学的研究结果及心理治疗的效果也存在类似的情况。然而，一旦心理学研究结果和心理治疗效果不能在所有情况下都适用，就常常会引起人们对心理学产生极大的失望和轻蔑。一旦面对心理学的话题，人们常常忘记一个最基本的原则，那就是知识不需要完全确定后才是有用的——即便某些知识不能预测个体的具体情况，但如果能对群体的总体趋势有预测能力，也是非常有益的。基于群体的特征所做的结果预测常常被称为总体统计数字或统计预测（下一章将详细讨论统计预测这一概念）。

人们经常为心理学预测设定一个比其他科学更高的标准。想想看，

当一个不健康的人去看病，医生说除非他进行锻炼和改变饮食习惯，否则有很高的风险发作心脏病。我们不会因为医生没有告诉这个人“如果不改变饮食习惯，他将于 2012 年 9 月 18 日心脏病发作”，而认为医生的信息是无用的。我们容易理解该医生的预测是概率性的，并不能达到那种精度。同样，当地质学家告诉我们，某地区在未来 30 年发生一场震级为 8.0 或更大地震的可能性为 80%时，我们不会因为他们没有说“2012 年 7 月 5 日就会有地震发生在这里”而贬低其知识。

然而，心理学却往往被设置了更高的标准。当学校心理学家推荐一个针对学习障碍儿童的训练计划时，显然是在做概率预测——该训练能使这些儿童有较大的可能性获得好成绩。当一个临床心理学家推荐一个针对有自我伤害行为的孩子的计划时，情况也与之类似。心理学家判断如果按计划进行治疗，会有较高的概率获得一个很好的结果。但是不同于心脏病发作和地震的例子，心理学家常常要面对诸如“但我的孩子何时能达到某一年级的阅读水平？”或“他在这个治疗计划中要待多久？”这类问题。这些问题都是无法回答的，正如地震和心脏病何时发生也是无法回答的一样，因为针对所有这些问题——心脏病发作、学习障碍儿童、地震以及自我伤害的儿童——所做的预测都是概率性的。

出于这些原因，全面认识概率推理对理解心理学至关重要。耐人寻味而又颇具讽刺意味的是，心理学很可能是人们不能进行统计思维的最大受害者，然而心理学家却是对人类概率推理能力研究最多的人。

有关概率推理的心理学研究

过去的 20 年里，普林斯顿大学的丹尼尔·卡尼曼（Daniel Kahneman，2002 年诺贝尔奖得主，见 MacCoun，2002）、密歇根大学的理查德·尼斯贝特（Richard Nisbett）及已故的阿莫斯·特维斯基（Amos Tversky）等心理学家的研究，彻底改变了我们对人类推理能力的认识。他们在研究中发现，很多人头脑里压根儿没有概率推理的基本原则，更

多人则是有一些但并不完备。正如学者经常指出的，这些基本原则在人们头脑里没有充分发展并不足为奇。作为数学的一个分支，统计学是最近才发展起来的（Hacking，1975）。而在概率定律被发现之前，机遇游戏已经存在了好几个世纪了。这又是一个例证：个人经验不足以让人们获得对世界的基本理解（参见第 7 章）。针对概率定律的正式研究发现了机遇游戏的运作机制，而成千上万的赌徒以及他们的个人经验，并不足以揭示机遇游戏的本质。

问题在于，社会越复杂，人们就越需要概率思维。如果一个普通人想要对生活其中的社会有一个基本的理解，那么，他至少应具备统计思维这一最基本的能力。

你或许有以下疑问：“为什么他们要提高我的保险费？为什么张三的保费比李四高，是不是社保局穷疯了？我们州的彩票有黑幕吗？犯罪率到底是在增加还是在减少？为什么医生要安排这些检查？为什么欧洲人可以用一些很珍稀的药，而美国人就不行？做相同的工作，女性赚的真的比男性少吗？国际贸易真的减少了美国人的就业机会，并降低了他们的薪酬吗？日本的教育要比我们好吗？加拿大的卫生保健真的比美国好且价格低廉吗？”这些问题都问得很好，这都是关于我们的社会如何运作的具体而实际的问题。要知道每个问题的答案，我们就必须运用统计思维。

显然，本书由于篇幅所限，不可能全面讨论统计思维。然而，我们将简要地讨论某些概率推理中的普遍误区。学习概率思维技巧的最好方法就是察觉人们在统计推理时最常犯的错误是什么。此外，对某些误区的了解对理解心理学发现及理论的重要性至关重要。

对概率信息的不充分利用

在心理学领域中，有一个已经被反复证实的发现，那就是一个具体事件的信息往往可以完全击败较为抽象的概率信息（第 4 章中讨论的“鲜活性”问题）。忽视概率信息的例子比比皆是，而且并不仅仅局限于缺

乏科学知识的外行人。卡斯塞尔斯、谢诺博格和格瑞博维斯（Casscells, Schoenberger, & Graboys, 1978）在哈佛医学院的四所教学医院中进行了一项研究，他们向 20 位医学专业的学生、20 位内科主治医师和 20 位办公室工作人员提出下面一系列问题："如果在每 1 000 人中有 1 人携带艾滋病病毒（HIV），再假设有一种检查可以百分百地诊断出真正携带该病毒的人；最后，假设这个检查有 5% 的阳性误诊率。也就是说，这项检查在没有携带 HIV 的人中，也会错误地检测出有 5% 的人是病毒携带者。假设我们随便找一个人来进行这项检查，结果呈阳性反应，表明此人为 HIV 携带者。假定我们不知道这个人的患病史，那么他真的是 HIV 携带者的概率是多少呢？

普遍的回答是 95%，正确的答案是约 2%。医生们过分高估了阳性结果表示患病的概率，因为他们一方面过分重视个案信息，另一方面又忽视了基础比率信息，从而过高地估计了阳性检测结果所真正代表的患病概率。稍稍进行逻辑推理就可以说明基础比率对概率的重要作用。1000 个人当中只有 1 人是真正的 HIV 阳性者。如果另外 999 人（不患病）也进行了此项检查，由于这一检查有 5% 的虚报率，他们当中将有接近 50 人（999 乘以 0.05）会被检查出携带这种病毒。这样一来，呈阳性反应的人就会是 51 个。因为在这 51 个人当中，只有 1 人是真正的 HIV 阳性者，此人确诊得病的概率其实只接近 2%。简而言之，基础比率就是绝大多数人没有携带这种病毒（病毒携带者只有千分之一）。这个事实和确定的虚报率综合考虑，就能使人确信，在绝对数量上，大部分呈阳性反应的人并不携带这种病毒。

尽管参与卡斯塞尔斯等人研究的医生们很快就意识到了以上概率逻辑的正确性，但他们最初的直觉反应却是忽视基础比率，并过分看重临床检测的证据。简单来说，事实上医生们知道什么是对的，但却本能地做出了错误结论。心理学家把这类问题称为*认知错觉*（参见 Kahneman & Frederick, 2002, 2005）。在认知错觉中，即使人们知道正确答案，他们也会由于问题的问法不同而做出错误的结论。

我们这里提到的所有例子都是认知错觉，因为它们都利用了人类推

理的误区：过分倚重个别事件所提供的证据而忽视了统计学的信息。对大多数人来讲，个案证据（实验室的研究结果）好像是摸得着的、具体的，而概率证据则好像是摸不着、不确定的。当然，这种理解是错误的，因为个案证据本身一定是概率性的。一项临床检验会以一定的概率对疾病做出误诊。上述情境就是一个例子，要想做出正确的决策，就必须结合考虑两种概率——对个案证据做出正确或错误诊断的概率（即 95% 或 5%）和过去经验所提供的*先验概率*（也叫*基础比率*）。整合这些概率的方法有的是正确的，也有的是错误的，并且时常是错的——特别是当个案证据给人一种很具体的错觉时（请回忆在第 4 章所讨论的鲜活性问题）——人们往往会以错误的方式来整合信息。这种概率推理的失败会极大阻碍心理学知识的应用，因为心理学的知识经常采用概率的形式来表述行为之间的关系。

科普作家科尔（K. C. Cole，1998）让我们想象下面两种情况。一种是用吸烟的死亡率来劝人不要吸烟，比如吸烟的死亡率是 0.000055，这是一种最常见的劝人方式。第二种方法则更为生动一些，让吸烟者想象在每 18 250 包烟中有一包是与众不同的——它里面装满了炸药，当吸烟者打开它时就会被炸死。我们绝对知道哪一个效果更好——然而它们表达的却是一个同样的事实。

样本大小信息的误用

请大家思考下面两个由特维斯基和卡尼曼（Tversky & Kahneman，1974）提出的问题：

1. 一个小镇里有大小两所医院。在大医院里每天大约有 45 个婴儿出生，在小医院里每天大约有 15 个婴儿出生。如你所知，大约有 50% 的婴儿是男孩，但具体的百分比每天都不一样，有时候高于 50%，有时候低于 50%。每一所医院都记录了一年内出生的男婴比例高于 60% 的天数。你认为哪一所医院记录的天数多？

a. 大医院

b. 小医院

c. 基本一样

2. 假设一个容器里装满了球，其中有 2/3 是一种颜色，其余 1/3 是另一种颜色。一个人从中拿出 5 个球，发现有 4 个是红色的，1 个是白色的。另一个人从里面拿出 20 个球，发现有 12 个是红色的，8 个是白色的。哪一个人会更自信地认为这个容器里有 2/3 的球是红色的、1/3 的球是白色的，而不是有 1/3 的球是红色的、2/3 的球是白色的？这两个人会给出什么样的概率呢？

对于第一个问题，大多数人回答"基本一样"。剩下的人则一半选择大医院，一半选择小医院。但正确的答案是小医院，所以接近 75% 的被试都给出了错误答案。答错是由于人们没有认识到，样本的大小在这个问题中的重要性。当其他因素保持不变时，较大的样本总是能够更精确地估计出总体的真正数值。也就是说，在任何一个指定的日子，较大的医院由于有较大的样本，男婴出生的概率更趋近于 50%。相反，小的样本总是倾向于距离总体平均值比较远。因此，小医院将会有更多的天数记录了与总体平均值相矛盾的男婴比率（60%，40%，80% 等等）。

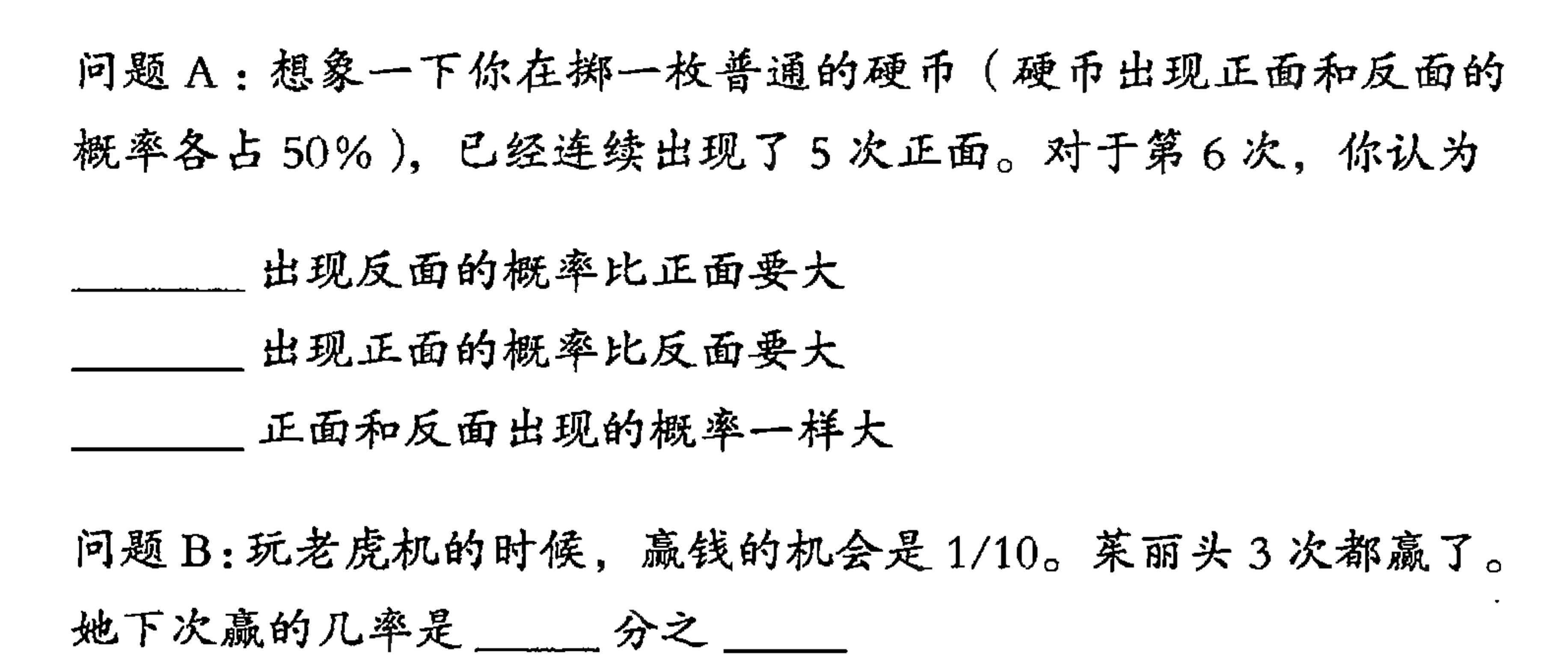
在回答第二个问题时，大多数人认为 5 个球的样本提供了更令人信服的证据，能证明这个容器里的球大多数是红色的。事实上，概率恰恰与之相反。对 5 球样本来说，坛里大部分为红球的几率是 8：1。而在 20 个球的样本中，这个几率是 16：1。尽管在 5 个球的样本中，抓出红球的比例较高（80%：60%），但考虑一下，另一个样本的大小是其 4 倍，因此对球的比例能够做出更为精确的估计。然而大部分被试被 5 个球的样本中红球有较高的比例给迷惑了，而没有充分考虑到 20 个球的样本具有更大的可信度。

在不同领域中进行证据评估时需要遵守的一条基本原则，就是认识到样本规模对信息可信度的影响，这对于理解行为科学的研究结果尤为重要。不管我们是否意识到，我们会对较大的群体持有一些普遍的看法。

我们很少察觉到，我们最坚定的信念是建立在多么脆弱的事实基础之上。把对几个邻居和同事的观察，以及在电视新闻上看到的一些趣闻轶事放在一起，我们就迫不及待地要对“人性”或者“美国人”发表见解。

赌徒谬误

请回答下面两个问题：



问题 A：想象一下你在掷一枚普通的硬币（硬币出现正面和反面的概率各占 50%），已经连续出现了 5 次正面。对于第 6 次，你认为

________ 出现反面的概率比正面要大
________ 出现正面的概率比反面要大
________ 正面和反面出现的概率一样大

问题 B：玩老虎机的时候，赢钱的机会是 1/10。莱丽头 3 次都赢了。她下次赢的几率是 ______ 分之 ______

这两个问题是为了检测你是否容易出现所谓的赌徒谬误——即倾向于将过去事件和未来事件之间联系起来，而实际上两者是独立的。两个结果是相互独立的，一个事件的出现不会影响另一事件出现的概率。大多数机遇游戏都具备这种性质。例如，幸运轮盘的数字与之前的数字无关。轮盘数字一半是红的，另一半是黑色的（为简化起见，我们将忽略绿色的零和双零），所以对任意一次旋转来说，出现红色的概率均等（0.50）。然而在连续 5~6 次出现红色数字之后，许多投注者转投黑色，因为他们认为现在黑色更有可能出现。这就是赌徒谬误：明明是独立事件，却认为先前的结果会影响下一结果出现的概率。在这种情况下，投注者错在他们的信念。轮盘并不记得先前发生过什么。即使连续出现 15 个红色数字，红色数字在下轮出现的概率仍然是 0.50。

在问题 A 中，有些人认为在 5 次出现正面之后，反面更可能出现。他们这么想就陷入了赌徒谬误。正确的答案是，正面和反面在第 6 次中

出现的可能性一样大。同样，对问题 B 任何非 1/10 的回答都落入了赌徒谬误。

赌徒谬误不仅限于没有经验的赌徒。研究表明，即使是那些一周赌 20 小时的资深赌徒，仍然表现出赌徒谬误（Petry，2005；Wagenaar，1988）。事实上，研究表明，正在接受赌博脱瘾治疗的个体比对照组更相信赌徒谬误（Toplak et al.，in press）。

重要的是我们要认识到，这一谬误不仅限于赌博游戏，它还存在于任何概率起着重要作用的地方。换句话说，它几乎存在于一切事情之中。婴儿的基因构成就是一个例子。心理学家、医生和婚姻顾问常常遇到一些已有两个女孩的夫妇，他们正计划要生第三个孩子，因为“我们想要个男孩，这回一定是个男孩”。这就是赌徒谬误，在生了两个女孩之后生男孩的概率（接近 50%）和生第一个孩子时完全一样。生了两个女孩不会增加第三个孩子是男孩的概率。

赌徒谬误存在于任何一个有几率成分的地方，如体育比赛和股票市场。一些心理学家（Gilovich，Vallone，& Tversky，1985；Burns，2004）研究了在篮球运动中对“连投连中”或“手热”的迷信，这一迷信是指，相信某一个投手能够变得“手热”，并且在连续投中之后，下一次投中的概率也会更高（“把球传给他，他现在手热”）。研究者证实，篮球运动员和球迷都十分相信“连投连中”。例如，在一个问卷调查中，91% 的篮球迷认为刚投中两球或三球的球员，与刚有两次或三次失误的球员相比，在下一次投篮时会有较高的投中概率；84% 的球迷认为，把球传给刚刚连续投中两球或三球的球员是重要的。当请球迷估计，假设一个球员在场地上有 50% 的投中率，那么在他投中或没投中一球之后，下一投投中的概率是多少。结果，球迷们对前者的估计是 61%，后者是 42%。研究者调查了费城 76 人篮球队的队员，结果发现大多数（但不是全部）球员对连投连中所持有的信念与球迷们几乎一样强烈（见 Gilovich et al.，1985）。

但是为什么我们要在赌徒谬误的标题下讨论连投连中呢？因为根本就没有连投连中这回事！吉洛维奇等人（Gilovich et al.，1985）研究了

费城76人队和波士顿凯尔特人队在1980~1981赛季中投篮命中的统计数据。在这一赛季，球员们的投篮并没有出现前后关联的现象。让我们从非技术的角度看看这代表了什么意思。

赌徒谬误相信独立事件间是有关联的，即认为毫无关联的事件之间存在依从关系。从统计学意义上，连投连中可以被解释为连续投中两球或三球后，投篮的命中率会高于前面几次未投中时再投的命中率。吉洛维奇等人（1985）计算了这个概率，发现没有任何证据支持这个假设。例如，朱利叶斯·艾尔文（费城76人队投篮次数最多的球员）的数据资料表明，他在连续三次投中后，接下来投篮的命中率为0.48，而连续三次未中接下来的命中率为0.52；在连续两次投中后，接下来命中率为0.52，而连续两次未中后，接下来命中率为0.51；在一次投中后，接下来命中率为0.53，在一次未中后，接下来命中率为0.51。简单来说，无论前几次投篮的情况如何，艾尔文的命中率都是接近0.50——压根儿没有连投连中这种事。

其他球员的资料也非常相似。莱昂内尔·霍林斯连续两次投中后接下来的投篮命中率是0.46，连续两次未中后，接下来的投篮命中率是0.49。他投中一次后，接下来的投篮命中率是0.46，和一次未中后接下来的命中率完全一样。这说明，不管霍林斯前几次投篮的结果如何，他投篮的命中率总是接近47%。波士顿凯尔特人队的罚球资料也说明了同样的情况。例如，拉里·伯德在投中一次罚球后下一次罚球命中的概率是88%，而一次罚球不中后，下一次罚球命中的概率是91%。纳特·阿奇巴德在投中一次罚球后，下次罚球投中的概率是83%，而一次不中后，下次罚球投中的概率是82%。由此可见，在罚球中也不存在连投连中。相信球员可以变得“手热”的信念确实是赌徒谬误的一个例子，也就是说，相信事实上独立的、毫无关系的事件间存在着联系。

有趣的是，赌徒谬误看起来是第6章讨论的“直觉物理学”——仅凭经验是无法告诉人们世界的真相的——的一个例证。吉洛维奇等人（1985）测试了大学篮球队员在空场地（即没有任何防守者）上练习在15码处投篮时的表现。他们让这些球员对100次投篮的命中率打赌。

队员肯定应该能赢，因为他们一般在这个距离上能够投中的概率约为50%，而且打赌的规则是，当球员投中时赢的要比没投中时输的多一些。然而，球员可以在每一次投篮前选择押多（这样赢得多，输得也多）或者押少（这样赢得少，输得也少）。显然，如果球员能够预测自己的成绩的话，就会赢得比较多。也就是说，当他们认为投中的概率高时，他们就会选择多下注；而当他们认为投中的概率低时，就会选择少下注。实验结果表明，就算是专业的球员也没有发现“手热”的现象：一次或多次投中后，再投中的概率并不比一次没中后再投时更高。然而，球员们却都认为存在类似“手热”的情况。他们在投中一球后，对下一次投篮所下的赌注，要高于在一次没投中后所下的赌注。结果证明，球员们根本不能预测自己的表现：他们预测的结果并不比随机水平好。

赌徒谬误来源于对概率的诸多错误认识。其中一个错误认识就是，如果一个过程真正是随机的，就不可能出现重复同一结果或某种模式的序列，哪怕是一个不起眼的随机事件（例如，掷 6 次硬币）。人们习惯性地低估了重复（正正正正）或某种模式（正正反反正正反反正正反反）在一个随机序列中出现的可能性。正因为如此，人们在模拟一组真正的随机序列时，常常适得其反地产生出一个很少出现重复和某种模式的排列。这是因为，人们往往会错误地让可能的结果尽量轮流出现，以为这样才称得上是随机抽样，这无疑破坏了真正的随机排列中可能出现的结构（Nickerson，2002；Towse & Neil，1998）。

那些声称自己有通灵能力的人可以轻而易举地利用人们的这一错觉。大学心理学课上常会进行这样一种演示，老师让一名学生准备 200 个数字的排列，这 200 个数字从 1、2、3 这三个数字中随机重复抽取。完成之后，不要让老师看到。接下来，让这名学生全神贯注于他写的第一个数字上，老师则来猜这个数字是什么。当老师说出他的猜测之后，这个学生再向全班同学及老师公布正确的答案。有人记录猜对的次数，直至猜完这 200 个数字。在实验开始之前，这个老师声称有通灵能力，可以在实验过程中用读心术来证明“通灵能力”的存在。通常在展示之前，老师会先问班里的学生，他猜测的成绩要达到多少——也就是“猜

中”的百分比是多少——才算是能证明他确实有通灵能力。这时，通常都会有一个修过统计课程的学生回答说，因为纯粹随机的猜测也能猜中33%，所以要想让别人相信他有通灵术，猜中的比例就一定要超过33%，至少达到40%。班上大部分同学都会认同这一个观点。演示结束后，结果那位老师猜中的比例果真超过了40%。这个结果令很多同学感到惊讶。

学生们从这一演示中领教了什么是随机性，并且知道伪装通灵能力是多么地容易。在这个例子中，老师仅仅利用了“人们不让连续重复的数字出现”这一事实：人们频繁地在三个数字间换来换去以制造“随机性”。在真正的随机序列中，已经出现了三个2之后，再出现2的概率是多少呢？其实还是1/3，与出现1或3的概率一样大。但大多数人在产生随机数字时并非如此。出现一个哪怕很小的重复片断之后，人们也常常会刻意地变换数字，力图制造一个“随机”序列。这样，在我们的这个例子中，老师只要在每一轮猜测前，不去挑选那个学生在前一轮中挑选的那个数字，而从另外两个数字中选一个就可以了。例如，如果那个实验中的学生在上一轮说的数字是2，那么老师就会在下一轮的猜测中从1或3中任选一个。如果学生在上一轮说的数字是3，那么老师就会在下一轮的猜测中从1或2中任选一个。这样一个简单的把戏根本不需要什么通灵能力，就能保证猜中的概率高于33%——高于三个数字随机猜测的准确率。

人们总是认为，如果一个序列是随机的，那它就不应呈现有重复和某种模式。2005年关于iPod（美国苹果公司出品的数码音乐播放器——译者注）“shuffle”模式（意即“随机播放”）的争议（Levy，2005）就以一种幽默的方式证明了这一点。此模式将下载到iPod里的歌曲以随机的方式播放。很多用户抱怨说shuffle模式并不随机，因为他们经常听到同一专辑或流派的歌曲。当然，许多心理学家和统计学家在听到这类抱怨时只能暗自苦笑，因为他们了解我刚才提到的类似研究。科普作家史蒂芬·列维（Steven Levy，2005）讲述了他经历过的类似事情。他的播放器似乎在起初的一个小时里偏爱史提利·丹（Steely Dan）的歌！但列维明智地接受了专家告诉他的事实：真正的随机序列，往往看起来

不像是随机的，因为我们倾向于给所有事物都套上一种模式。在进行有关问题的研究后，列维总结道，“生命可能确实是随机的，iPod 可能也是。但是，我们人类将永远有自己的套路和模式，只为让无序变得可控。即使真的存在缺陷，问题也不在 shuffle，而在我们自己身上”（p. 10）。

再谈统计与概率

以上列举的涉及统计推理理解中出现的错误，仅为冰山一角，有可能阻碍人们正确理解心理学。有兴趣的读者可以阅读由吉洛维奇（Gilovich）、格里芬（Griffin）和卡尼曼 (Kahneman) 编写的《思维捷径和偏见：直觉判断心理学》（*Heruistics and Biases: The Psychology of Intuitive Judgment*，2002），它在这一方面提供了比较完整、详细的描述。

吉格瑞泽（Gigerenzer）的《计算的风险：如何察觉数字是在欺骗你》（*Calculated Risks: How to Know When Numbers Deceive You*，2002）对统计与概率做了很通俗的介绍（对没有受过任何数学训练的初学者尤其适用）。此外还有哈斯戴（Hastie）和达维（Dawe）的《不确定世界的理性选择》（*Rational Choice in an Uncertain World*，2001）和拜农（Baron）的《思考和抉择》（*Thinking and Deciding*，2000）以及尼克尔森（Nickersn）的《认知和几率：概率推理的心理学》（*Cognition and Chance: The Psychology of Probabilistic Reasoning*，2004）。

本章中所讨论的概率思维具有巨大的实践意义。由于没有充分运用概率思维能力，医生们选择了效果欠佳的治疗方法（Baron，1998；Dawes，2001）；人们不能准确地评估环境风险（Margolis，1996）；在法律程序中错误地使用信息（Foster & Huber，1999；Lees-Haley，1997）；政府和私人企业将数以百万计的资金用于不必要的项目（Arkes & Ayton，1999）；动物不断被捕杀以至濒临灭绝（Baron，1998；Dawkins，1998）；对病人实施了不必要的手术（Dawes，1988，pp.73–75）；有人做出了错误的财务判断，损失巨大（Belsky & Gilovich，1999；Kahneman & Tversky，2000；Zweig，2001）。

当然，我们不可能在一个章节里全面地讨论统计推理。我们的目的就是想强调统计对于研究及理解心理学的重要性。不幸的是，当遇到统计信息时，我们还找不到一个放诸四海皆准的规则。功能化的推理技能不像科学思维中的其他部分那么容易获得，而是需要通过正规学习才能掌握。幸运的是，现在大多数综合大学和社区学院都提供了入门级的统计学课程，而且不需要大学程度的数学基础。在上这类课程之前，读者可以先阅读我刚才推荐的那些书。

尽管很多科学家都真诚地希望一般大众能够知悉和理解科学知识，但有时对一门学科的精通依赖于对某些信息的掌握，而对这些信息的掌握又只有通过正规的学习才能实现。如果说对一门学科的深入理解是一般外行人也能随便达到的，这是一种在学术上不负责任的态度。统计学和心理学就属于这一类学科。心理学家艾伦·班欧（Alan Boneau，1990）调查了心理学教科书的作者，请他们列出学生在学习心理学时需要掌握的最重要的术语和概念。在所列出的术语及概念中，频率最高的100个词汇中有接近40%属于统计学和方法论的范畴。不精通统计和概率的人不可能成为称职的心理学家（Evans，2005；Friedrich，Buday，& Kerr，2000）。

不可否认，本书的一个目的就是要使心理学的研究能为广大读者所接受。然而，心理学进行理论建构所依靠的实证方法和技术与统计学是如此密不可分（这一点和其他很多领域一样，如经济学、社会学和遗传学），以至于没有一个人可以在对统计学毫无知晓的情况下精通心理学。因此，尽管这一章对于统计思维介绍得相当粗略，但它的主要目的是要凸显另外一个对于理解心理学至关重要的专业领域。

小　结

和大多数学科一样，心理学研究所得出的是概率式的结论——大多数情况下会发生，但并非任何情况下都发生。虽然这些结论并非是

100% 准确的（就像其他科学中的情况一样），但根据心理学研究及理论所做出的预测仍然是有用的。阻碍人们理解心理学研究的一个原因就是，人们很难用概率的术语来思考。在这一章里，我们讨论了几个相当精彩的研究实例，这些例子表明大多数人如何与概率推理背道而驰：当人们遇到具体的、具有鲜活性的证据时，就把概率信息抛到一边了。他们没有考虑到，较大的样本能够提供对于总体数值更为精确的估计。最后，人们表现出赌徒谬误（把原本无关的事件看成是有联系的）。赌徒谬误源于下一章将要讨论的一个更为普遍的倾向：未能认识到偶然性在决定结果时所起的作用。

11

偶然性在心理学中扮演的角色

在上一章里我们讨论了概率趋势、概率思维和统计推理的重要性。本章将沿袭这一话题，重点强调人们理解随机性和偶然性这两个概念时遇到的问题。我们将强调，由于没有领会偶然性是如何始终贯穿于心理学理论中的，人们常常误解了研究对于临床实践的贡献。

试图解释偶然性事件的倾向

我们大脑的进化始终以这样一种方式，就是让我们能够不懈地寻求世界中的各种模式。我们寻求身边事物的关系、解释及其背后的意义。心理学家已经对这种强烈倾向进行了研究。这是人类智力的典型特征，而且能够解释人类在信息加工和知识获得过程中表现出来的令人惊叹的诸多能力。

然而，人类认知过程的这种极具生存适应性有时也会反戈一击。例如，环境中没有什么可以进行概念化的东西，可我们还是还一味地去寻求概念性的理解，这就是一种不良适应。那么，到底是什么在人类认知这一最与众不同的方面制造麻烦呢？是什么打乱了我们对结构的寻求并

阻碍了我们对事物的理解呢？你猜对了，是概率。说得更具体些，是偶然性和随机性。

偶然性和随机性是我们周围环境不可分割的一部分。偶然性和随机性的规律支配着生物进化和基因重组的机制，物理学也运用关于偶然性的统计定律来解释物质的基本结构。自然界发生的很多事情，都是系统性以及可解释的因素与偶然因素共同作用的结果。再回想一下前面谈到的例子：吸烟导致肺癌。生物学上系统的、可解释的方面将吸烟和某一疾病联系起来，但这并不表示所有吸烟者都会患肺癌，这种趋势是概率性的。或许最终我们能解释为什么有些吸烟者不会患肺癌，但在现阶段，这种变异性必须归因于大量偶然性因素，是这些因素决定一个人是否患某一疾病。

这个例子说明，当一件事取决于偶然性时，并不一定表示它是*不确定的*，只是说它目前是*无法确定的*。掷硬币是偶然事件，但并不是说在对抛掷的角度、硬币内的金属成分以及许多其他变量加以测量之后，也不可能确定其抛掷的结果。实际上，这些变量确实决定了掷硬币的结果。但是，我们称掷硬币为随机事件，是因为在每一次抛掷时，我们没有比较简易快捷的方法来测量这些变量。一次抛掷的结果并不是严格意义上的不确定，它只是在当下无法确定而已。

世界上的许多事件不能以系统性的因素来完全解释清楚，至少现在还不能。然而，当一个特定的现象没有现成的系统解释的时候，我们头脑中的概念寻求“设备”往往仍在隆隆运转，试图将无意义的理论强加于原本随机的数据。心理学家曾对此现象进行了实验研究。一个实验情境中，要求被试观察一系列在多个维度上有所区别的刺激物，并告诉他们其中的一些刺激物属于一类，而其他的则属于另一类，被试的任务是去判断每一个刺激物属于这两类中的哪一类。实际上，刺激物是研究者随机归类的，因此除了随机性，并没有任何其他规律。但是，被试很少敢做随机猜测。相反地，他们通常会绞尽脑汁，虚构一个复杂的分类方法，并对他们给这些刺激物所做的分类做出解释。

同样地，形形色色的“预谋论”通常也需要一套又一套复杂的说辞

去解释那些预谋论者拼命想要理解的、原本是由随机因素导致的事件。这一现象十分典型，甚至各类权威人士在其专业领域内的工作也往往如是。许多金融分析师的思维方式就体现了这一谬误。他们通常会对股票市场价格的每一次小的波动都编造出精细的解释，而实际上这种变化大多只是随机波动而已（Malkiel，2004；Taleb，2001）。然而，股票市场分析师总是不断对客户暗示他们可以（也许他们也相信自己可以）“征服市场”，即使当大量的证据表明他们中的大部分其实是做不到这一点的。过去几十年中，如果你购买了标准普尔指数中的所有 500 种股票，然后放着不去管它（我们称之为“傻子策略”的办法——去买一种依照这一指数的互惠基金），那么今天你获得的回报会比 2/3 的华尔街股票经纪人为他们的顾客所赚的还要高（Egan，2005；Hulbert，2006；Malkiel，2004；Updegrave，1995），你的成绩也会打败 80% 订阅费已经涨至每年 500 美元的财经通讯杂志（Kim，1994）。

但是，我们要如何看待那些确实打败了傻子策略的经纪人呢？你可能想知道这是否意味着他们具有某些特殊的才能。我们通过设想这样一个实验来回答这个问题：有 100 只猴子，每只猴子手中握有 10 支飞镖，它们都向一面写有标准普尔 500 指数的墙上掷飞镖，飞镖扎中的股票就是那年要买的股票。那么，一年后它们的业绩会是怎么样的呢？有多少只猴子能打败标准普尔 500 指数？恭喜你答对了。大概有一半的猴子会。那么，你会不会愿意付钱给这一半打败标准普尔 500 指数的猴子，授权它们在下一年帮你选股呢？

这个关于财经预测的例子的延伸，证明了原本纯粹随机的事件会因怎样的逻辑而看起来像是由可以预测的因素造成的（Fridson，1993；Paulos，1988）。假想你收到一封信，信中告诉你有这样一份关于股票市场预测的通讯。这个通讯并不收费，只是要求你试试照着他们的建议去买股票，然后看看它的预测灵不灵。它告诉你 IBM 的股票会在下个月攀升。你把这份通讯随手一扔，但是你确实注意到在下一个月里 IBM 股票果真涨了。如果你曾读过一本与本书的内容类似的书，你会觉得这是稀松平常的事情，仅会将其视为一次侥幸的猜中。后来你又收到另一

份来自同一家投资咨询公司的通讯，该通讯说 IBM 股票会在下个月下跌，当股票确实跌了的时候，你仍将其视为侥幸，但是这一次你可能就有点儿好奇了。当这家公司寄来第三份通讯，预测 IBM 下个月会再次下跌时，你发现自己对这几页财经内容的关注度提高了。继而你发现该通讯又一次做出了准确预测，IBM 这个月确实又下跌了。当来自这家公司的第四份通讯说 IBM 下月会涨，而且也确实涨了时，你难免会觉得这个通讯真还挺神，而情不自禁地想花 29.95 美元去订一年这本如此有价值的通讯。这种诱惑难以抵挡，除非你能想象：此时在一个简陋的地下室里，某人正在准备下周要寄出的 1600 份通讯，这些通讯会按电话黄页上的 1600 个地址发出，其中 800 份预测 IBM 下月上涨，800 份预测下跌。当 IBM 在下个月真的涨了，公司就继续把通讯只发给上月接收到正确预测的 800 位“客户”（当然，其中还是 400 份预测涨，另外 400 份预测跌）。然后，你可以想象，这个“锅炉房”——可能还包括在背后煽风点火、辅助造势的电话营销骗子——正在向第二周接收到正确预测的 400 位客户发送第三个月的预测通讯（还是 200 份预测涨，200 份预测跌）。是的，你就是连续四次收到正确的随机预测信息的 100 个幸运儿之一！这 100 个“幸运儿”中的大多数会为了能继续收到通讯而支付 29.95 美元。

现在看来这就像是一个玩弄众人于股掌之上的可怕骗局。实际也是如此。而当那些“受人尊敬”的财经杂志或电视节目给你推荐“连续四年击败一半以上对手的股票经纪人”时，情况也好不到哪儿去。请回想一下猴子掷飞镖的场景，设想这些猴子是年年选股的股票经纪人。很明显，第一年他们之中有 50% 会击败他们的对手。第二年，这 50% 的人中又有一半——按随机水平来说——会击败其对手，即 25% 的经纪人能连续两年击败他们的对手。之后第三年又有一半——随机水平——能击败对手，即总人数的 12.5% 连续三年击败对手。最终到第四年，又会有这些人的一半（总人数的 6.25%）能击败自己的对手。因此，100 只猴子中大概有 6 只能取得像财经节目和报纸所说的“连续四年击败了其他的经纪人”的骄人成绩。那么，这 6 只击败了一起扔飞镖的同伴的

猴子（正如你所见，也击败了大多数现实生活中的华尔街经纪人；参见 Egan，2005；Malkiel，2004）的确有资格在电视节目“华尔街一周”中亮相，你觉得呢？

解释偶然性：错觉相关和控制错觉

人们有解释偶然事件的倾向，这一现象在心理学的研究中称为错觉相关。当人们相信两类事件在通常情况下应该一起发生时，就会认为自己频繁地看到了同时发生的现象，甚至当这两类事件的同时出现是随机的，并不比任何其他两个事件同时发生的频率更高时也是如此。总之，即使是面对随机事件，人们也倾向于看到他们所期望的联系（Nisbett & Ross，1980；Stanovich，1999，2004）。他们在原本没有规律的地方看到了规律。

许多有控制的研究（如 King & Koehler，2000；Stanovich & West，1998）都证明，当人们头脑中已经预设了两个变量相互关联的想法时，他们甚至能够在两个变量根本毫无关系的数据中发现联系。不幸的是，这一发现在现实生活中也广泛存在，并对人们的生活产生负面影响。例如，许多从事心理治疗工作的人一直都对罗夏墨迹测验的效度深信不疑。这个著名的墨迹测验要求被试对一张白纸上的墨迹做出反应。因为这一墨迹缺乏结构，所以其理论是，人们会以自己对模糊情境的典型反应来对这些墨迹做出反应，从而揭示其“潜藏的”心理特质。这种测验也被称为投射测验，因为它假定被试会将他们潜意识的内心活动和感受投射在墨迹上。然而问题是，没有任何证据表明当罗夏测验作为一个投射测验而使用时，提供了任何额外的诊断价值（Garb，Florio，& Grove，1998；Lilienfeld，1999；Lilienfeld et al.，2000；Wood，Nezworski，& Stejskal，1996；Wood，Nezworski，Lilienfeld，& Garb，2003）。对罗夏测验的信心是源自于错觉相关这一现象。临床心理医生从病人的反应模式中看到了联系，是因为他们相信本来就有这种联系，而不是真的从反应模式中观察到了什么联系。

心理学家雷·海曼（Ray Hyman）讨论了人们喜欢在原本没有模式地方的寻找模式的倾向：

> 我们不得不运用自己原有的知识和期望以获得对世界万物的理解。在大多数的一般情境中，这种对于知识背景和记忆的运用让我们正确地阐释一些主张，并对此提供必要的推论。但这一强有力的机制在原本没有承载任何信息的情境中偏离正轨。有些本是能够轻而易举就识别出来的随机性噪音，我们却不懈地要从中寻找意义（Hyman，1981，p.96）。

在我们的生活中，许多人际交往里都包含大量的偶然成分："互不相识的男女约会最终促成了婚姻；取消约谈而丢了工作；误了班车而遇到了高中的老同学，等等。认为生活中每一件偶然的小事都需要精细的解释，这种想法是错误的。但是，当偶然事件确实会产生重要的后果时，人们不免要建构一些复杂的理论去解释它们。

试图去解释偶然事件的倾向可能源于我们深切地渴望相信自己是可以控制这些事件的。心理学家埃伦·兰格（Ellen Langer）研究了控制错觉这一现象，指的是人们有一种倾向，愿意相信个人能力可以影响偶然事件的结果。在一项研究中，两个不同公司的雇员向同事兜售彩票，一些人只是简单地塞到手里，而另一些人则可以自行抽取。当然，在随机抽奖的事件中，彩票是自行抽取的还是派发的没有什么区别，中奖率都是一样的。但是，第二天，当这两个雇员试图向这些同事买回彩票时，自行抽取彩票的被试对彩票的要价是被派发者的四倍！在另外几项实验研究中，兰格证实了这一假设，之所以出现这些结果，是因为人们不能接受个人因素无法影响偶然事件这一事实。这一错觉广泛存在的证据来自于美国各州发行彩票的经验。这些州充斥着教人们如何"征服"彩票的伪科学书籍。这类书之所以畅销，是因为人们不懂得随机性的含义。事实上，自从 20 世纪 70 年代中期新泽西州发明了一种新的彩票售卖方式之后，美国各州才爆发购买彩票的热潮。这种方式就是让购买者可以自行刮奖或自行挑选号码（Clotfelter & Cook，1989；Thaler，1992，

p.138）。用这种售卖方式来进行的抽奖活动通常叫做“参与性抽彩”，而这类参与性抽彩正是利用了当时兰格研究的控制错觉现象：人们错误地相信他们的参与行为能够决定随机事件。

还有一些心理学家则研究了另一个与此相关的现象，该现象被称为公平世界假设，它是指人们倾向于相信自己是生活在一个公平的世界里，在这里每个人都得到他们应得的东西（Hafer & Begue，2005）。研究者发现了一些实验证据，证明了公平世界中存在一种“罪有应得”的信念：人们会鄙视那些偶然不幸的受害者。为偶然事件寻求解释的倾向导致了这一现象。人们很难相信一个完美无瑕的或是道德修养高的人会因为偶然事件而惨遭不幸。固然我们想要相信好人有好报、恶人有恶报，但是，偶然性是不偏不倚的，它以完全不同的方式运行：好事坏事都以相同的概率发生在不同人身上。

当公平世界假设这一信念被推向极致时，会催生出一些非常有害或是不人道的教条。且看20世纪80年代早期美国教育部一名官员的逻辑，他说，残疾人“错误地认为生命中的巧合令他们偶然受到惩罚，实际并非如此。所有发生在某个人成长过程中的某个点上的事，没有一件不是由他本人所招致的……这句话听上去好像很不公平，但是一个人所处的外部环境确实是与他内在心灵的发展相一致的”（Gilovich，1991，p.143）。正如吉洛维奇所指出的，“这真不该是一个想进入教育部高层的官员所持有的哲学，教育部本应是负责给予残疾人士同等教育机会的机构”（pp. 143~144）——但如果我们拒绝将这类后果归为偶然性，结果必然会导致这种不人道的哲学。

公平世界假设中所体现的对于偶然性的错误理解，也助长了其他一些错误的民间信念，导致人们容易看到虚假相关。例如，我们在第6章中提到过，“盲人有非常敏锐的听觉”就是一个错误的信念，这个错误信念可能会一直流传下去，因为这种联系能体现“老天很公平”，而这正是人们希望看到的。

偶然性和心理学

在心理学中也存在这样的倾向：研究者试图解释一切，希望其理论不仅能解释行为中系统的、非随机的成分，还要能解释任何细微的变异。这种倾向导致了不可证伪的心理学理论的泛滥，既包括个人提出的理论，也包括那些看似科学的理论。“心理历史学”的奉行者常常犯下此类错误。一个著名人物生命中的每一个细小的变化及转折，都经由精神分析学派的理论在心理历史中得以诠释。大多数心理历史事件存在的问题是，不是它们解释得太少，而是它们解释得太多。这一研究方法的奉行者很少承认一个人的一生是由许多偶然因素决定的。

对于想要运用心理学知识的外行人来说，理解偶然性这一因素的作用是非常重要的。受过正规训练的心理学家承认他们的理论只能解释人类行为变化的一部分而非全部，他们会坦然面对偶然因素。但是，那个在奥普拉秀中出现的（见第 4 章开头）能对每一个个案及人类行为的每个细节做出解释的嘉宾，引发的不是崇拜而是质疑。真正的科学家从不惧怕承认自己的无知。总之，评价心理学主张的另一实用法则就是：在接受对某个事件的复杂解释之前，先想一想偶然因素在其中扮演了什么角色。

巧 合

为纯粹偶然的事件寻求解释的这种倾向，也导致我们对许多巧合事件的性质产生误解。许多人认为巧合需要特别的解释，他们不理解巧合的发生并不需要偶然性之外的因素，巧合并不需要特别的解释。

《韦氏新世界字典》（*Webster's New World Dictionary*）里把巧合定义为：“相互关联或相同的事件意外地、令人不可思议地同时出现了。”鉴于这本字典把意外定义为“偶然地出现”，所以这个定义不存在问题。巧合只是相关事件偶然地同时出现。不幸的是，许多人并不这样解释巧

合。那些在事件中寻求模式和意义的倾向与巧合“不可思议的”的特性结合在一起，让许多人忘记他们可以用偶然这一因素来解释巧合，反而为理解这一现象寻求特别的解释。下面讲的这个故事你一定已经听过无数次了：“那天我正坐在那儿寻思，我好久没给德克萨斯州的老比尔叔叔打电话了，紧接着电话铃就响了，你猜怎么着！正是我那老比尔叔叔打来的。这种心灵感应的背后肯定有点儿什么原因！”这就是一个典型的为巧合事件编造解释的例子。每天，我们大多数人都可能想到很多或远或近的人，这些人在我们想起他们时，有多少人可能会打电话来呢？几乎没有可能。这样一年之内，我们可能想过数百个不曾打来电话的人。最终，在经历数百次这种我们不曾意识到的“错误尝试”之后，某个人在我们想他/她的时候正准备给我们打电话。这种事情难得一见，但难得一见的事情也会发生——纯粹是偶然。其他解释都是画蛇添足。

如果人们真正理解了巧合的含义（一个偶然发生的令人不可思议的事件），他们就不会落入陷阱去寻求系统的、非偶然性的解释。但事实正相反，对很多人来说，巧合是需要偶然性以外的原因来解释的。例如，许多人都听到过这样的说法：“天哪！简直太巧了！我真想知道为什么！”为此，马科斯（Marks，2001）建议大家今后用罕见匹配这个比较中性的名词来形容令我们感到惊异的两个事件的同时出现。

有一种错误信念助长了为巧合事件寻求解释的倾向，这种信念认为罕见的事不会发生，罕见匹配也绝非偶然。我们的这类错误信念之所以如此强烈，是因为概率有时是用几率（odds）这一词语来表述的，而这种表述具有双关的暗示作用（英语中“odds”既有“几率”的含义，还有“罕见”的含义——译者注）。看看我们是用什么方式来表述概率的：“啊！天哪，这事儿是极不可能发生的！因为它出现的几率只有 1/100!”我们在做这样的表述时所用的方式让人强烈地感觉到这件事绝不会发生。当然，我们可以用另外一种表达方式来表述同一件事，而这一方式可能给人带来完全不同的感受：“在 100 个同类事件中，这种结果可能会出现一次。”这种表述方式强调，尽管这一事件是少见的，但是长时间来看，罕见的事终究一定会发生的。简言之，罕见匹配是会偶然发

生的。

事实上，概率定律确保了随着事件发生次数的增加，一些罕见匹配出现的可能性会变得很大。这一定律不仅允许罕见匹配出现，而且从长远来看几乎保证了它的出现。请看马科斯（Marks，2001）的例子，如果一次掷 5 枚硬币，结果它们都是正面朝上，你将认为这是一个罕见匹配，一件不太可能的事情。是的，它发生的概率是 1/32 或 0.03。但是如果你将这 5 枚硬币掷 100 次，再问，在这 100 次中，至少有一次全部正面朝上的可能性是多少呢？答案是 0.96，就是说，100 次中，这一罕见匹配是极有可能发生的。

若干年前，安·兰德炮制了一系列流传甚广的有关亚伯拉罕·林肯总统和约翰·肯尼迪总统之间令人“毛骨悚然”的巧合：

1. 林肯于 1860 年当选总统；肯尼迪则于 1960 年当选。
2. 林肯和肯尼迪都关注民权。
3. 林肯和肯尼迪这两个名字都有 7 个字母。
4. 林肯有一个秘书叫肯尼迪，肯尼迪也有一个秘书叫林肯。
5. 两人都由叫约翰逊的南方人继任。
6. 两人都被有三个名字的人暗杀（John Wilkes Booth 和 Lee Harvey Oswald）。
7. Booth 和 Oswald 都持有不受人欢迎的政见。
8. Booth 在剧院里射杀了林肯，然后藏在仓库里；Oswald 从仓库中射杀了肯尼迪，然后藏在剧院里。

当然，作为巧合，这些事之间的联系一点也不令人毛骨悚然。德克萨斯大学的电脑程序师约翰·李维（John Leavy，1992）曾经搞了一个“令人脊背发凉的总统巧合竞赛”来说明，实际上在任何两个总统之间找到像上述那样的一个清单是多么容易的事（见 Dudley，1998）。例如，李维的文章中比较了威廉·亨利·哈里森和扎卡里·泰勒、波尔克和卡特、加菲尔德和麦金利、林肯和杰克逊、尼克松和杰斐逊、华盛顿和艾森豪威尔、格兰特和尼克松、麦迪逊和威尔逊之间的相似之处。下面是加菲

尔德和麦金利之间惊人的相似之处：

1. 麦金利和加菲尔德都生长在俄亥俄州。
2. 麦金利和加菲尔德都是美国内战的老兵。
3. 麦金利和加菲尔德都在众议院任过职。
4. 麦金利和加菲尔德为了保护美国工业，都支持金本位制和关税保护制。
5. 麦金利和加菲尔德这两个名字都有 8 个字母。
6. 麦金利和加菲尔德都被来自纽约市的副总统取代：西奥多·罗斯福和切斯特·亚伦·阿瑟。
7. 罗斯福和阿瑟两人的名字都有 17 个字母。
8. 两个副总统都蓄须。
9. 麦金利和加菲尔德都在任期内第一年的 9 月被枪杀。
10. 刺杀他们的凶手，查尔斯·基埃图和利昂·乔尔戈什听起来都不像美国人的名字。

许多关于总统之间联系的清单都很相似。总之，考虑到一个人几十年的生命中人际交往和各类事件的复杂性，在这样一个包含成千上万个事件的样本空间里，任何两个人之间如果找不到什么相似之处才是让人感到奇怪的（Martin，1998）。

懂得在什么时候避免对纯粹随机因素导致的事件编造复杂的解释，这是具有实际作用的。作家艾图尔·嘉万迪曾描述了 1973 年赎罪日战争期间，认知心理学家卡尼曼和以色列空军打交道的事例。两个飞行中队出发并返航，一队损失了四架飞机，另一队则没有损失。军方希望卡尼曼调查一下，之所以有这样的差异，是否有特别的因素在起作用。卡尼曼并没有去做调查，他仅仅运用了本章所谈到的理念去告诉以色列空军不要浪费时间：“卡尼曼知道，如果空军官员真的去调查，他们将不可避免地在两个中队间发现一些可测量的差异，并且感觉非得做点什么”（Gawande，1999，p. 37）。但是卡尼曼知道，任何找到的因素都极有可能是虚假的——不过是纯粹的偶然性波动的结果而已。

个人的巧合

发生在我们个人生活中的罕见匹配往往对我们具有特殊的意义，我们尤其不愿将其归因为偶然。产生这种倾向的原因有很多，某些是动机性和情感性的，还有一些是概率推理的失败。我们通常不能意识到，罕见匹配只是巨大“概率事件”样本库中一个非常小的部分而已。对我们中的某些人来说，罕见匹配看起来好像经常发生，但是它真的经常发生吗？

想想如果我们现在对你个人生活中的罕见匹配加以分析，会得到什么结果。假定某一天里你参与了 100 件不同的事情。考虑到现代工业社会中生活的复杂性，这个数字并没有高估，实际上可能还低估了。你看电视、打电话、与人面谈、讨论去工作或去商场的路线、做烦人的家务、看书获取信息、在上班时完成复杂的任务等等。所有这些事件都包含很多可单独记忆的成分。这样一算，100 件事其实真不算多，不过，我们就按100件事情来算。罕见匹配是指其中两个事件不可思议地联系在一起了。那么典型的一天中这 100 件事之间共有多少不同的、两两匹配的组合呢？用一个简单的公式就能算出结果，你通常一天有 4 950 个不同的配对组合，而一年有 365 天。我们知道，罕见匹配是令人难忘的，比尔叔叔打来电话的那一天可能令你数年难忘。假如你把 10 年内所记得的所有罕见匹配数出来，也许也就 6 或 7 件（或多或少，人们对于小概率有不同的标准）。这 6、7 件事情来自于一个多大的概率事件样本库呢？每天 4 950 个配对事件，乘以一年 365 天，再乘以 10 年，得到 18 067 500 个配对。总之，10 年中如果有 6 个你认为是罕见匹配的联系发生了，就有 18 067 494 个也可能是罕见匹配的其他配对事件发生了。所以，你的生活中的一个罕见匹配发生的概率是 0.00000033。有 6 个罕见匹配出现在 1800 万个事件中，的确很稀罕，但并不奇怪。罕见的事件确实发生了，它们也的确少见，但是，偶然性这一因素保证了它们一定会发生（回忆前面掷 5 枚硬币的例子）。在我们的例子中，6 件奇事发生在你身上，它们可能是巧合：两个相关事件由于偶然性的存在而不可思议地同时发生了。

心理学家、统计学家以及其他科学家都指出，许多罕见匹配实际上并没有人们通常认为的那么“罕见”。著名的“生日问题”是最好的例子。在一个23人的班级里，有两个人生日是同一天的概率是多少？大多数人会认为非常低。而实际上，23人的班级中，两人同一天过生日的可能性大于50%。而在35人的班级，可能性就更大了（概率大于0.80；见Martin，1998）。所以，因为美国历史上有43位总统，因此詹姆斯·波尔克和沃伦·哈丁两位在同一天出生（11月2日）也就不足为奇了。同样地，有38位总统都已过世，其中米勒德·菲尔莫尔和威廉·塔夫脱死于同一天（3月8日）也不应令人感到惊讶，甚至还有另外3位总统——约翰·亚当斯、托马斯·杰斐逊、詹姆斯·门罗——都死于同一天，而这一天竟然是7月4日，美国独立日！后面这个神奇吗？其实不过是概率使然罢了。

接受错误以减少错误：临床预测与统计预测

在试图解释世界上发生的所有事，同时又拒绝承认偶然因素的作用，实际上会降低我们对现实世界的预测能力。在某个领域中，承认偶然因素的作用意味着研究者必须接受这样一个事实，即我们的预测不可能百分之百准确，预测中总是会犯一些错误。但有趣的是，承认我们的预测达不到百分之百的准确度，实际上反而有助于我们提高整体预测的精确性。这听起来好像有点儿矛盾，但是事实确是如此：为了减少错误就必须接受错误（Dawes，1991；Einhorn，1986）。

“我们必须接受错误以减少错误”这一概念可以通过一个在认知心理学实验室里研究了数十年的非常简单的实验任务来证明（Fantino & Esfandiari，2002；Gal & Baron，1996）。这个实验任务是这样的，被试坐在两盏灯（一红一蓝）前，实验者要求他们去预测每次测试时哪一盏灯会亮，被试要参与很多轮这样的测试，并按准确率给予一定的报酬。实际上，所有的测试都是在70%的次数亮红灯、30%的次数亮蓝灯的

条件下进行的，两种灯以随机顺序出现。实验过程中，被试很快就感到红灯亮的次数比较多，因此也就在更多的测试中预测红灯会亮。事实上，他们确实在大约 70% 的测试中预测红灯会亮。然而，正如前面所讨论的，被试在实验过程中逐渐发现并相信灯亮是有一定模式的，但却从没想过序列是随机的。为了要使他们的预测百发百中，他们在红灯与蓝灯之间换来换去，保持 70% 的次数预测红灯会亮，30% 预测蓝灯会亮。被试极少意识到，如果他们放弃对“百发百中”的追求，他们的预测会更好一些！为什么会是这样的呢？

让我们想想这一情境背后的逻辑。在以 70∶30 的比例随机点亮红灯或蓝灯的情况下，如果被试在 70% 的测试中预测红灯会亮，30% 的测试中预测蓝灯会亮，他的准确率会是多少呢？我们将用实验中间部分的 100 个测试来计算——因为那时被试已经注意到红灯亮的次数比蓝灯多，从而开始在 70% 的测试中预测红灯会亮了。在 100 次测试中有 70 次红灯亮了，所以被试在这 70 次中有 70% 的正确率（因为被试在 70% 的测试中预测红灯会亮），也就是说，被试在 70 次中有 49 次正确的预测；100 次测试中有 30 次蓝灯亮了，被试在这 30 次中有 30% 的正确率（因为被试在 30% 的测试中预测蓝灯会亮），也就是说，被试在 30 次中有 9 次正确的预测。因而，在 100 次测试中，被试的正确预测是 58 次。但是，请注意，这是多么可怜的成绩啊！如果被试在注意到哪一盏灯亮得比较多后，就总是预测那盏灯会亮——在本实验中，就是注意到红灯亮的次数比较多，因此就总是预测红灯会亮（姑且称之为“百分百红灯策略”），那么，他在 100 次测试中会有 70 次正确的预测。虽然在蓝灯亮的 30 次测试里，被试将没有一次正确的预测，但是总准确率仍然高达 70%——比在红灯与蓝灯之间来回变换以追求“百发百中”的 58% 的准确率要高 12 个百分点！

然而，百分百红灯策略取得的高准确率是要付出代价的：必须放弃“百发百中”的愿望。（显然，蓝灯偶尔亮的时候，被试始终是在预测红灯亮，也就是放弃了在蓝灯亮的测试中命中的机会）。这就是接受错误以减少错误。放弃不犯错误的想法，会让被试获得更高的总体准确度。

同理，以一定的精度预测人类的行为时，有时也需要接受错误以减少错误，也就是，在依靠一般性的原则来做出比较准确的预测的同时，也要承认我们不可能在每件具体事情上都对。

但是，“接受错误以减少错误”做起来很难。在心理学领域里，40年来关于临床预测和统计预测的研究就证明了这一点。统计预测是指依据统计资料中得出的群体趋势所作的预测。本章一开始所讨论的群体（也就是总体）预测就是属于这种预测。一种简单的统计预测是，针对凡是具有某种特征的所有个体，做出相同的预测。例如，预测不吸烟者的寿命是 77.5 岁，而吸烟的人是 64.3 岁，就是一个统计预测。如果考虑的群体特征不只一个（运用第 5 章谈到的复杂相关技术——尤其是多元回归技术）将令我们的预测更加准确。例如，预测吸烟、肥胖且不运动者的寿命是 58.2 岁，就是在一个多变量（吸烟行为、体重和运动量）基础上的统计预测，这样的预测总是比单变量的预测更加准确。统计预测在经济学、人力资源、犯罪学、商业与市场学以及医学等领域都很常见。

在心理学的许多分支领域，如认知心理学、发展心理学、组织心理学、人格心理学与社会心理学中，其知识都是通过统计预测来表述的。相反，一些临床心理从业者则声称他们可以超越群体预测，对特定个体做出百分之百准确的预测，这种预测被称为临床预测或个案预测。与统计预测相反，临床预测是这样的：

> 某些心理学家声称，他们能对个体进行预测，从而超越了对“一般人”或不同类别的人所进行的预测……某些心理学家最大的不同在于，他们主张将每个人理解为独一无二的个体而不是群体的一部分，而统计概括是适用于群体的。某些心理学家声称能分析出在个体的生活中“什么导致了什么”，而不说“总体而言”什么是对的。（Dawes，1994，pp.79–80）

临床预测似乎可以视为是对统计预测的有用补充，但问题是，临床预测并不准确。

如果证明临床预测是有效的，那么一个临床医生与他的病人接触的经验以及有效运用病人所提供的信息，应该使他能够提出比较好的预测，这个预测一定能胜过对病人信息进行编码、然后输入能够对量化数据加工的统计程序而得到的预测结果。总之，有人主张说，临床心理从业者的经验使得他们能够超越尚未由研究揭示的关系。"临床预测是有效的"这一观点很容易验证，不幸的是，经过检验，这一观点被证明是错误的。

对临床预测与统计预测的比较研究所得到的结果始终是一致的。自从保罗·米尔（Paul Meehl）的经典著作《临床预测与统计预测》（*Clinical Versus Statistical Prediction*）于 1954 年出版以来，40 年间有超过 100 个研究表明，在几乎每一个曾经验证过的临床预测领域（精神治疗的效果、假释行为、大学生毕业比例、电击治疗的反应、累犯问题、精神病住院治疗期的长短等等），统计预测都优于临床预测（Dawes，Faust，& Meehl，1989；Faust，Hart，Guilmette，& Arkes，1988；Goldberg，1959，1968，1991；Ruscio，2002；Swets et al.，2000；Tetlock，2005）。

在多个临床领域中，研究者给临床心理医生一份病人的信息，让其预测这个病人的行为。与此同时，他们也把同样的信息加以量化，用一个统计方程加以分析，这一方程是以先前研究发现的统计关系为基础编制的。结果都是统计方程大获全胜。这就表明，统计预测比临床预测更为准确。事实上，即使是在临床心理医生可以获得比统计方法更多的资料的情况下，后者仍然比前者的预测更准确。也就是说，临床心理医生除了拥有与统计预测一样的量化资料以外，还拥有与病人单独接触和访谈所得到的资料，但是这并没有令其预测变得像统计预测那样准确。"即使拥有信息优势，临床判断仍然不能超越统计方法；实际上，拥有更多的信息，并不能弥补两种方法之间的差距"（Dawes et al.，1989，p. 1670）。产生这种结果的原因当然是统计方程将各种信息数据按照优化标准整合起来，并且做得准确而稳定。优化和稳定这两个因素就让临床心理医生通过非正式方法收集到的资料和信息的优势消失殆尽。

检验临床－统计预测的研究文献中，还包含这么一种方法，那就是

给临床心理医生由统计方程得来的预测结果，让其根据自己与病人接触的经验来对这一预测做出调整。结果，临床医生对统计预测做出调整后，预测的准确度非但没有增加，反而降低了（见 Dawes，1994）。在这里我们又看到了一个不能“接受错误以减少错误”的绝好例子，与前面所述的那个红蓝灯预测实验非常类似。应当利用灯亮次数多少这一统计信息而采用每次都预测红灯的策略（可以获得 70% 的正确率）时，被试却为追求次次正确而在红灯与蓝灯之间换来换去，结果正确率反而降低了 12%（只有 58% 的次数是正确的）。同样地，在上述研究中，临床心理医生相信，他们的经验应该可以提供给自己一些“洞察力”，从而得以做出比定量数据更好的预测。实际上，这些“洞察力”根本不存在，他们的预测比依赖公开的统计信息所做出的预测要差。最后需要指出的是，统计预测的优越性并不局限于心理学，它业已扩展到了许多其他临床科学中——例如，医学中对心电图的解读（Gawande，1998）。

对于研究显示统计预测优于临床预测的优势，米尔（Meehl，1986）曾说：“社会科学中，没有任何一个争议能如这次这般，从这么大量的、性质上如此多样的研究中得到如此一致的结论。”（pp. 373–374）。但令人尴尬的是，心理学领域并没有应用这一知识。例如，这个学科在研究生入学与心理健康培训招生等程序中仍然不停地使用个人面试，尽管大量证据表明，面试方法缺乏效度。临床工作者也继续利用一些似是而非的证据来证明他们对于“临床直觉”的依赖是合理的，而不依靠更有效的总体性预测。例如，道斯等（Dawes et al.，1989）曾指出：

> 一种普遍的反统计论调或误区在于，认为群体统计不适用单个人或事。这种观点是对概率基本原则的误用……要保持逻辑上的一致，反统计论的鼓吹者就必须相信并承认，如果一个人被迫玩一次俄罗斯轮盘赌，允许他选择膛内装有 1 发或 5 发子弹。事件的单一性使得选哪把枪都无所谓（p. 1672）。

关于这一点的一个类比是，问你自己对如下科学发现的反应是什么，这个发现是：完成过多次类似手术的医生，在下一例手术中成功的概率

会比较高（Christensen，1999）。现在有一个医生 A，他常做某一类手术，失败的可能性很小，而另一个医生 B 从没做过这种手术，失败率可能很高，请问，你愿意让这两个医生中的哪一个来为你做手术呢？如果你相信“概率不适用于个案”，那你就不该介意让医生 B 给你做手术。

在诸如心理治疗效果等问题上，承认统计预测优于临床预测并不会对心理学的声望造成任何损失，因为在医学、商学、犯罪学、会计学甚至是家畜鉴定等许多领域中，这条规律都适用（见 Dawes，1994；Dawes et al.，1989；Dowie & Elstein，1988）。尽管从总体上说，心理学不会因为这些研究结果而有什么损失，但是对那些以“专家”身份出入各种活动，并让病人相信他们有独一无二的临床个案知识的临床心理从业者来说，当然会造成声誉或者收入上的损失。然而，正如麦佛和瑞特（McFall & Treat，1999）在一篇论述临床评估价值的文章所提醒的那样：“我们试图评估和预测的事情在本质上是概率性的。这意味着我们不能期望大自然会如此听话，能让我们以百分百的把握去预测单一事件。相反，我们最高的期望也只能是鉴别一系列可能的结果，然后去估计每个结果出现的相对可能性。从这种概率的角度看，传统临床评估期望达到的那种理想化目标——对独特的未来事件做出精确的预测——其实太天真了，反映了我们的无知或自大，或二者兼有”（p. 217）。

实际上，如果我们将“接受错误以减少错误”变为一种习惯，心理学和整个社会都将从中受益。在试图对每一个不同寻常的事件做出独特解释时（就我们目前的知识情况来说，独特的解释也许根本不可能），我们常常丧失了对更多平常事件的预测能力。请大家再次回想一下红灯－蓝灯实验，诚然，“百分百红灯策略”会对出现概率较小或很少出现的不寻常事件（蓝灯亮）做出错误的预测，但如果我们把注意力放在出现概率较小的事件上，采用“70% 红灯、30% 蓝灯策略”，结果会怎样呢？我们会在 30 个不寻常事件中正确预测 9 次（30 × 0.3），其代价是丧失了对 21 个常见事件做出正确预测的机会，没有对红灯做出 70 次的正确预测，只获得 49 次的正确预测（70 × 0.70）。临床领域中的行为预测也遵循相同的逻辑，为每一个案编造复杂的解释，确实可能抓住一小部分

不寻常事件——但这是以损失了对大多数事件的正确预测为代价的，而在此方面，简单的统计预测则更有效。加望德（Gawande，1998）指出，医学领域也同样需要学习“接受错误以减少错误”这个道理。他认为在医学里，强调直觉、个别化的治疗方法“是有缺陷的——我们试图承认并考虑人类复杂性的因素，但这非但没有避免错误，反倒招致了更多的错误”（p.80）。

华格纳和科瑞（Wagenaar & Keren，1986）论证了对个人知识的过分自信以及对统计信息的忽视，会破坏“系安全带驾车”的交通安全推广活动的效果。因为人们总是认为：“我和别人不一样，我驾车很安全”。问题是85%的人都认为“自己的技术比一般驾车者高明”（Svenson，1981）——这显然是很荒谬的。

“统计数据不适用于单一个案”这一同样的谬误，是导致赌徒积习难改的重要因素。华格纳（1988）在他的赌博行为研究中总结道：

> 从我们和赌徒的讨论中可以非常清楚地看出，赌徒大体上都能意识到赌博造成的不良后果。他们也知道最终输的会比赢的多，而且在未来也是如此。但他们却不能把这些统计性的思路应用到下一局、下一小时或下一个晚上。丰富的直觉经验还是让他们觉得，统计学在下一局或下一小时里派不上用场，他们相信自己能够预测下一局的结果（p. 117）。

华格纳发现，强迫性赌徒对“接受错误以减少错误”有很强的排斥倾向。例如，二十一点牌局的玩家，普遍拒绝使用一种基本策略（见Wagenaar，1998，第2章），这种基本策略可以保证把庄家的胜率从6%或8%降低到不足1%。基本策略是一个长期性的统计策略，强迫性赌徒之所以拒绝它，是因为他们坚信“有效的策略应该是在每一把都有效”（p. 110）。华格纳研究中的赌徒“总一成不变地说，这类系统的一般性策略是不会有用的，因为它们忽略了每一个具体情境的独特性”（p. 110）。这些赌徒抛弃能保证他们少输上千美元的统计策略不用，转而去徒劳地追求建立在每一具体情境独特性基础之上的“临床预测”。

当然，这里有关临床 - 统计预测研究文献的讨论，并不意味着个案研究在心理学中毫无价值。请大家记住，这一章所谈的只是“对行为的预测”这一特定情境。回想一下在第 4 章中对于个案研究价值的讨论，个案信息在引发对重要的、需要进一步研究的变量的关注方面是非常有用的。而这一章中所说的则是，一旦相关的变量已经确定，我们要开始运用它们来预测行为时，测量这些变量并使用统计公式来进行预测始终是最优程序。首先，我们通过统计方法得到了更为准确的预测；其次，统计方式优于临床预测之处在于，统计程序所得出的预测是公共知识，任何人都可以使用、修改、批评或争论。相反，如果使用临床预测就等于要依靠个别权威的评估——由于这类判断太过个别和特殊——因此不能接受公众的评议。正如道斯（Dawes，1994）所述：

> 一个职业心理学家在法庭上宣称，他有一种效果极佳的“临床判断”，是基于他自己多年所积累的经验，无法公开证实，只能在诸如专业资历、做出相同证言的年数、态度品行等无关联的方面接受挑战。与之相反，一个统计模型可以在理性方面接受挑战，因为它是公众性的（p. 104）。

小　结

偶然性在心理学中扮演的角色时常被外行人士和临床心理从业者所误解。人们很难认识到，行为事件结果的变化中有一部分是由偶然因素造成的。也就是说，行为的变化有一部分是随机因素作用的结果，因此心理学家不应自诩能够预测每一例个案的行为。心理学的预测应该是概率性的——是对总体趋势的概率性预测。

表示自己可以在个体层次上进行心理预测，是临床心理学家常犯的错误。他们有时候会错误地暗示别人，临床训练赋予了他们一种对个别案例做出准确预测的“直觉”能力。恰恰相反，几十年来，有价值的研

究都一致表明：在解释人类行为的原因方面，统计预测（基于群体统计趋势的预测）远远优于临床预测。目前还没有证据表明，临床直觉能预测一个统计趋势是否会在一个特定的个案身上出现。因此，当对行为进行预测时，千万不要对统计资料置之不理。统计预测也昭示，当对人类的行为进行预测时，错误和不确定性将始终存在。

12

不招人待见的心理学

虽然公众对心理学话题怀有浓厚兴趣，但他们对于心理学及其所取得的成就给出许多负面评价。心理学家们都意识到了这个“形象问题”，但他们又感到无能为力，所以干脆不去管它，这样做其实是错误的。当大众传媒在决定公众感知（例如，虚构的电视“纪录片”对那些知识储备不足的公众来说就成了真实的历史）方面越来越有影响力的时候，不理会心理学的形象问题只会让情况变得更糟。

罗德尼·丹杰费尔德（Rodney Dangerfield）是 30 多年来非常走红的一位美国喜剧演员，他标志性的口头禅是：“我得不到尊重！”从某种意义上说，这也正是心理学在一般公众心目中的形象写照。本章就是想谈谈为什么心理学会像丹杰费尔德那样无法获得应有的尊重。

心理学的形象问题

之前我们曾讨论过造成心理学形象问题的成因。例如，在第 1 章中所讨论过的弗洛伊德问题，无疑导致了人们对心理学较低的评价。如果要公众列举一个著名心理学家的话，这个人不是弗洛伊德就是斯金纳。

对于其理论的歪曲报道在公众中广为传播，导致心理学被认为是一门非常肤浅的学问。当一门学科中最具影响力的学者被误解为主张人没有思想、主张人和老鼠没有差别的时候，这门学科还有什么希望？斯金纳当然没有否认人类能够思考（Gaynor，2004），他从动物身上发现的许多有关操作性条件反射的定律，已被证实的确能推广到人类行为上。然而，公众对这些科学事实却不甚了了。同样地，对弗洛伊德理论的歪曲也降低了心理学在公众心目中的形象。

心理学和超自然心理学

除了弗洛伊德和斯金纳的研究之外，外行人对其他卓越的心理学研究几乎一无所知。想证明这一点，到附近的书店去看看公众能买到什么样的心理学读物就知道了。你的调查会发现，那些摆在书店卖的心理学读物通常可以分为三类。第一类是心理学的一些早期经典著作（弗洛伊德、斯金纳、荣格、弗洛姆、埃里克森等），这些著作多半侧重老式的精神分析观点，已经完全不能代表当代心理学了。令心理学家感到沮丧的是，这一领域最有价值的著作都被淹没在书店的科学或生物学类书籍中。例如，心理学家史蒂芬·平克尔（Steven Pinker）的名作《思维的运作》（*How the Mind Works*）总是被归在科学类而非心理学类图书中。因此，他所探讨的关于认知科学的重要文章，被迫与生物学、神经生理学或计算机科学而非心理学为伍。

在多数书店中可以找到的第二类读物，是那些伪装成心理学的伪科学书籍，里面充斥着无数超自然现象，如心灵感应、千里眼、意念移物、超前感知、转世、生物节律、星灵投射、金字塔力量、植物沟通、通灵手术等（Lilienfeld，Lohr，& Moirer，2001）。书店里的心理类书籍中这类货色大量存在，无疑导致——也反映了——人们的误解：心理学家就是证实这些超常现象存在的人。这种误解对心理学而言多少有点儿讽刺。事实上，心理学与那些超常现象之间的关系很容易说清楚。这些超常现象压根不在现代心理学感兴趣的范畴之内。个中缘由可能会令许多

人大吃一惊。

超自然体验和其他超常能力的研究不被认为是心理学的一部分，此观点可能会激怒许多读者。多个调查结果都显示，超过 40% 的公众相信超感知现象的存在，并且狂热地信奉自己的信念（Farha & Steward, 2006；Musella，2005；Rice，2003）。历史研究和调查指出了公众热衷于这类信念的原因（Alcock，1987；Humphrey，1996；Lilienfeld，2005；Stanovich，1989，2004）。像大多数宗教一样，许多所谓的超常现象也标榜诸如转世之类的说法。对部分人来说，来世的说法能满足其超越现有生命极限的需求。心理学研究“不识时务”地指出超自然体验得不到证实，无疑粉碎了这些人的热切企盼。心理学主张不把超感知视为一个可行的研究领域，不可避免地会引发其信徒的不满，他们控诉说，心理学家把这一类主题排除于心理学研究之外的做法是独断专行的。如果心理学家仅是摆出无可奈何的姿态并无视这些反对的声音，这将无助于增进公众的理解。与之相反，心理学家应该针对这些反对观点的谬误给予细致而清晰的解释。这样的解释要强调科学家们并不是根据什么法令来确定研究主题的，也没有什么条例指出什么能研究、什么不能研究。研究领域的兴起、延续或终结，所依据的是理论及方法的自然选择过程。那些产生出丰硕的理论及实证发现的领域都获得了大量科学家的认可；而那些理论上行不通或者没有能被重复验证的领域就会被摒弃。这种对理论与方法的自然选择引导着科学向真理靠近。

例如，超感官知觉在当代心理学中不被认为是一个可行的研究主题，就是因为其研究一直无法积累任何正向的成果，所以它让大部分心理学家失去了兴趣。在这里我要强调“当代”一词，是因为多年以前心理学家确实对超感官知觉怀有极大的兴趣，直到累积了大量的负面证据之后，这种兴趣才消退了。正如历史所展示的那样，研究课题通常不是由某个权威政府机构宣布停止的，它们只是在生存竞争的环境中被自然淘汰出局了而已。

在心理学领域里，超感官知觉从来没有被认为是一个不能研究的主题，这一事实是清楚并且公开的（Alcock，1990；Hines 2003；

Humphrey，1996；Hyman，1992，1996；Kelly，2005；Marks，2001；Milton & Wiseman，1999）。有许多研究超感官知觉的论文发表在专业的心理学刊物上。那些在媒体上频频曝光的超自然心理学家，总喜欢让人们觉得这一领域是崭新的，惊人的新发现即将出现。其实，事实却没什么新鲜的。

对超感知的研究和当代心理学自身的历史一样久远，它并不是什么全新的研究领域。在心理学文献中，它也曾经像许多现在被认为是可行的主题一样被认真地研究过。然而，在正式心理学刊物上所发表的有关这一领域的许多研究结果都无法证明超感知存在。在上世纪 90 多年的研究之后，我们仍然无法在控制实验条件下重复验证任何超感知现象。尽管过去几十年来进行了大量有关超感知的研究，却从来没有一个研究能达到这一简单而基本的科学标准。这一点甚至连超自然心理学家及其信徒都承认（参见 Alcock，1990；Hines，2003；Druckman & Swets，1988；Krippner，1977）。简而言之，尚未出现需要科学解释的未经证实的现象（见 Alcock，1990；Hines，2003；Hyman，1992，1996；Milton & Wise-man，1999）。仅仅这个原因，就使得心理学对这一话题失去了兴趣。

颇具讽刺意味的是，心理学家在评估超常能力方面扮演着关键角色。他们的重要性可能仅次于那些拆穿无数超能力演示骗局的专业魔术师（Randi，1986，1987）。而且，很多论述和质疑超常能力的重要书籍都出自心理学家之手。

讽刺意味显而易见。心理学作为一门最可能精确评估超感知言论的学科，在公众的心目中却与伪科学关系最近。这种“被连累”的现象让心理学深受其害。正如下面还要再详细讨论的，心理学常常会陷入这样一个“里外不是人”的境地，这只是其中一例。那种认为在心理学里没有什么规则、这个领域的知识缺乏科学评判标准的信念，导致人们将心理学与超感知这样的伪科学联系在一起。然而，如果心理学家成功地让公众认识到这些伪科学的真面目，心理学与伪科学的联系又会被视为“心理学不是一门科学”的铁证！

自助类读物

书店里常见的第三类心理学读物就是所谓的自助类读物。当然，这类读物也有许多不同种类（见 Fried，1994，1998；Fried & Schultis，1995；Paul，2001；Santrock，Minnett，& Campell，1994）。有一些书是励志类的，目的是为了提升人们的自我价值感和自信心。另一些书则是新瓶装旧酒，将一些关于人类行为的老生常谈重新包装了一下。只有少数（简直是凤毛麟角）书籍是由负责任的心理学家为公众撰写的。还有许多书，虽出自心理学专业人士之手，但算不上“负责任”的作品。为了标榜其“独特性”，声称自己发明了一些新“疗法”，不但可以矫治某些特殊行为问题，而且还能满足老百姓的一般需要（赚钱、减肥和拥有更好的性生活是其“三大”主题），这类书籍常能大卖。这些所谓的新疗法很少基于控制实验的研究，如果作者是个临床医生的话，他们通常只是依靠他们的个人经验或者少数的几个病例，就提出了自己的“发现”。

许多有效性经过了严格的心理学检验程序验证的认知和行为疗法，却很少出现在书店的货架上。这一情况在电子媒体中更为糟糕，电台和电视台几乎没有任何正规的心理学报道，相反，他们总是邀请一些江湖术士和爱出风头的媒体名人，而这些人与真正的心理学毫无瓜葛。媒体之所以会这样做，主要原因是，正规的心理疗法从来都不会声称自己能立竿见影、药到病除，甚至不会担保治疗一定成功，或者夸大其治疗的范围（如，“你不仅会把烟戒掉，而且生活的方方面面都会得到改善！”）。同样，现在互联网上也出现了类似的情形——由于缺乏严格的同行评审，人们在网上看到的治疗方法通常都是骗局。

这类在美国图书市场中占相当比重的自助类读物，极大地影响了公众对于心理学的印象。首先，像弗洛伊德问题那样，这些书使公众搞不清心理学研究关注的焦点在哪里。举个例子来说，虽然有相当数量的心理学家在为肥胖、人际关系和性问题提供治疗，并在不断地进行研究，但这个数量比起自助类读物中所说的要少得多。这种误解也使得公众以

为大多数的心理学家都致力于异常行为的研究和治疗。事实上，大部分的心理学研究的是人类的正常行为。正如美国心理学会前任主席马丁·塞里格曼（Martin Seligman，2002）所说，对于公众来说，心理学“几乎成了治疗心理疾病的同义词，心理学应该从疾病治疗的错误轨道回归其历史使命，那就是让正常人的生活更加丰富和完满”（p. 19）。

除了引起对研究内容的误解之外，自助类读物还让人们对心理学的研究方法和目的产生错误印象。正如第 4 章中所讲的那样，心理科学并不认为几个个案研究、见证叙述和个人经验就能构成支持某种疗法有效性的充分的实证证据——而这些却恰恰是大多数自助类“疗法”安身立命的根本。自助读物因此误导了社会大众，使他们认为大多数的心理学理论就是基于此类证据得出的。在第 8 章中我们已阐明，证实一个理论需要许多不同类型的证据来支持，个案研究所提供的数据的说服力在其中是最弱的。将此类证据视为证实某一理论或疗法的确凿证据，无疑犯的是根本性的错误。

菜谱式知识

最后，自助类读物使公众误解了心理学的目标和多数心理学研究所追寻的知识。心理学家利·夏法（Leigh Shaffer）认为这种读物带给人们一种强烈的印象，那就是认为心理学研究者所追求的是那种“菜谱式”的知识（Shaffer，1981）。*菜谱式知识*是指那些只告诉你如何去使用某物，但对其基本的运作原理一概不谈的知识。例如，大多数人知道如何使用电话，他们知道如何拨号，如何获得信息，如何打长途等等。但许多人对电话操作背后的物理原理一无所知。他们不知道电话的通话功能是如何实现的，他们只是会用。这就是电话的菜谱式知识。在我们的社会里，许多有关科技产品的知识都是菜谱式知识。

当然，这也不完全是一件坏事。事实上，多数技术产品的设计初衷，就是为了让那些对其背后的运作原理一无所知的用户也能使用。事实上，菜谱式知识这一概念提供了一种方法，可以概括基础研究和应用研究之

间的区别。基础研究工作者寻找自然界的基本原理，而不去考虑这些原理能否转化为菜谱式知识。应用研究工作者则致力于将基本原理转化成一个个只需菜谱式知识就能使用的产品。

多数自助类读物只提供关于人类行为的菜谱式知识，它通常能够简化为这样的形式，“你只要做 X，你就会变得更加 Y 了”，或者 “做 Z，某 A 就会表现出更多的 B”。当然，如果这个药方是正确的（这一假设往往并不全然成立），这么做也不为过。许多正规的心理治疗都提供了大量菜谱式知识。然而，当人们错误地认为，所有心理学研究的终极目标就是提供菜谱式知识时，问题就产生了。尽管许多心理学研究者确实致力于将基本的行为理论转化为实用的心理疗法、保健行动方案或有效的工业组织模式，但心理学主要还是一门发现行为的普遍事实和理论的基础学科。这就是心理学研究为何会让外人觉得很怪的另一个原因：基础理论的研究与应用研究之间存在巨大的差异。

如果一个人走进分子生物学实验室并询问一位研究者：我们在头痛时是应当服用两片还是三片阿司匹林，我们会觉得这个人很傻。原因并不在于分子生物学与缓解疼痛没有任何关系，事实上对止痛药的研究可能会运用到这一领域的知识。我们之所以说这个问题问得傻，是因为分子生物学家并不是那种在开药方水平上工作、回答你是要吃两片还是三片阿司匹林的人。研究者所关注的是有关生物成分在分子水平上的基本数据。这些数据可能会为许多不同领域提供菜谱式知识，但发现基本数据和将这些数据转化为菜谱式知识的人不大可能是同一个人；而转化为菜谱式知识的方法，也会与最初发现事实的方法有所不同。

由于自助读物让公众错误地相信，多数心理学家都是开发菜谱式知识的，这使得许多心理学家所做的基础研究显得颇为奇怪。海奇特曾让被试在一间黑暗的房间里注视着一个小红灯，这到底与我们的现实世界有什么关系？是的，从表面上看来，确实没有一点儿关系。海奇特是对人们的视觉系统如何适应黑暗的基本原理感兴趣，这些基本原理最终会转化成能够用来应对具体问题的菜谱式知识，例如夜盲症是由维生素缺乏导致的。然而，这一转化并不是由海奇特本人来完成的，而且它在几

年之后才到来。

因此，自助读物给公众对心理学的感知带来两种不良的副作用。第一，这些读物中涉及的问题并不能代表当代心理学关注的焦点，相反，它们通常反映的只是消费者想看的内容。心理学学生往往无法充分地意识到，图书出版是一种商业行为，市场的力量决定了什么样的读物可以摆在书店的书架上。然而，科学的关注点并不是由此决定的。在所有的学科尤其是心理学中，科学家们认为，有生命力的想法和那些被包装后能够热销的想法，两者之间可谓是判若云泥。

最后，自助类读物让心理学看起来像是一种纯粹的菜谱式知识。尽管这并不全错，但它不能反映心理学领域所进行的大量基础研究。

心理学和其他学科

当然，心理学并没有垄断对于行为的研究。许多其他的相关学科采用不同的技术和理论视角，也对我们关于行为的知识有所贡献。许多涉及行为的问题都要求多学科的取向。然而，大多数心理学家必须要接受的一个非常残酷的事实就是，当这种多学科问题的研究成果发表时，心理学家的贡献往往会被其他学科所掩盖。

关于心理学家的贡献被忽略、抹杀或者被部分归为其他学科的例子不胜枚举。例如，第一个有关电视暴力对儿童行为影响的研究是由美国公共卫生局主持的，研究结果发现二者之间存在因果关系，因此，之后由美国医学会通过一项决议，重新确认该项研究的成果并向公众推广。这本来也是顺理成章的事，的确没什么错，但这一举措在无意间造成了一个后果，就是媒体不断地将电视暴力的研究成果与美国医学会联系在一起，给公众造成了这样一个印象，即确立这一发现的研究是由医学专业人士主持的。事实上，绝大多数有关电视暴力对儿童行为影响的研究都是由心理学家完成的。

另一个导致心理学家的工作经常被划入其他学科的原因是，这些

年来，“心理学家”一词的含义已经含混模糊了。许多心理学研究者在标识自己时，往往把自己的研究专长加在“心理学家”之前，例如自称生理心理学家、认知心理学家、工业心理学家、进化心理学家和神经心理学家。还有一些称谓甚至摒弃了“心理学家”一词，例如神经科学家、认知科学家、社会生物学家、人工智能专家和行为学家等等。所有的这些举动，再加上媒体认为“心理学不是一门科学”的偏见，都导致了心理学家的成就被误划入其他学科：生理心理学家的成果被划归生物学，认知心理学家的成果被归为计算机科学，工业心理学家的成果被归入工程学和商学，等等。即使当代最杰出的心理学研究者之一丹尼尔·卡尼曼（Daniel Kahneman）获得了 2002 年的诺贝尔经济学奖，心理学也没分享任何好处！当然，诺贝尔奖中没有为心理学单独设立奖项（Benjamin，2004；Kahneman，2003；MacCoun，2002）。

埃莫里大学耶克斯灵长类动物研究中心的负责人，心理学家弗雷德里克·金（Frederick King）曾讲到，某天他花了大段时间解释动物模型对人类神经障碍研究的重要性，在聆听完这位在癫痫症的神经和行为研究方面成就斐然的学者长时间讲解之后，有位记者问道：“你不过是个心理学家，怎么会知道这么多关于癫痫症的事儿呢？”（King，1993）。

在 20 世纪 70 年代末，法院判决了几起有关标准测验的诉讼案。其中的一起是 PASE 起诉汉农的案子，涉及智力测验中的文化偏见问题。审理该案的法官认为，能够帮助他裁定的惟一方法，就是由他来亲自检查每一道测验题，并且相信他自己的直觉。他对自己能做出正确判决的能力深信不疑，并对测验的每个问题写下自己的个人看法（Bersoff，1981，1982）。这位法官断定，在这些标准测验中，一组测验中的八个题目和另一组测验中的一个题目可能是带有偏见的。这个法官没有意识到，像这样的问题其实是可以通过科学方法来检验的实证性问题。个人意见不仅与真相无关，甚至可能是极具误导性的。决定这些标准测验题目是否带有偏见，需要复杂的统计程序和收集大量的数据。心理学家已经并继续为评估和收集必要数据及发展相应的统计技术方面做出了突出的贡献。

颇具讽刺意味的是，研究实际发现，对于哪些题目存在文化偏见的问题，外行人的直觉判断通常都是错误的。许多被认为没有偏见的题目实际上却存在着多种偏见，而许多表面上看来不公正的题目却没有发现统计意义上的偏见（Sandoval & Miille，1980）。举个例子来说，韦氏成人智力量表在加拿大受到批评，因为其中某一分量表（“信息”分量表）中的一些题目似乎有偏袒美国公民之嫌。例如，其中有一个题目要求被试写出1950年之后美国四位总统的名字。因此，当这个测验在加拿大使用时，一些题目就被“加拿大化”了（Violato，1984，1986）。例如，那道“总统”的题目被改为“1900年之后加拿大四位总理的名字”。然而，就算是这样一个显而易见的、“常识性”的改变也产生了一个小问题：加拿大公民在“总统版本”的智力测验得分比在“总理版本”中的得分要高。

我们是自己最坏的敌人

我们不是只会把心理学的形象问题怪罪在其他人头上，心理学家自己在这方面也“功不可没”。试图把真正的心理学介绍给公众的正规心理学家往往得不到什么好的回报。然而，美国心理学会（APA）和美国心理协会（APS）正致力于促进与公众的沟通，美国心理协会为此还新创办了一本期刊，名为《公众感兴趣的心理科学》（*Psychological Science in the Public Interest*）。心理学需要在这一方面再加把劲。不然的话，如果公众在将来还对这一学科心存误解，那就只能说我们是自作自受了。

美国心理学会前主席罗纳德·福克斯（Ronald Fox）在最近的致辞中谈到了心理学在沟通和传播方面的问题，以及我们自身是如何带来这些问题的：

> 一些经常在大众传媒上露脸的从业者，他们的做法是不专业的、不道德的，并且使他的同行蒙羞……我们的学科对于那些不负责任

的、令人发指的公开欺骗缺乏有效的对策……当今世界里，公众成天接触的观点和意见都来自于一些骗子（在最近的一个电视脱口秀节目里，一个心理学家声称他已经帮助许多病人回忆起前世所受的精神创伤），而不是理性的心理学从业者。（Fox，1996，pp. 779–780）

最后，心理学的某些分支中存在着一些反科学的态度和现象（Coan，1997；Watters & Ofshe，1999）。例如，在一些心理治疗的圈子里，有人一贯拒绝对自己所采用的疗法进行科学评估。专栏作家和心理治疗师查尔斯·柯瓦斯阿默（Charles Krauthammer）写了一篇文章，论述了这种态度对心理治疗的声誉造成的严重损害（Krauthammer，1985）。第一，由于拒绝去莠存良，造成各种疗法泛滥成灾。这种泛滥不仅使消费者的权益受到损害，而且还加深了这一领域的误区："心理治疗已经步入混乱状态，因为……它很少让流派消亡，它无法在内部实现有效的淘汰，心理治疗正在鱼龙混杂中走向灭亡。"柯瓦斯阿默在这里所悲叹的是，不遵循证伪原则的做法已经妨碍了这一领域的科学进步。

柯瓦斯阿默最后又指出了心理治疗这个圈子的一个内在矛盾，一方面，他们认为心理治疗"更像一门艺术而非科学"，因此反对以科学的方法进行评估；另一方面，他们仍然非常关注所谓的"800 磅大猩猩"，即政府补贴和个人健康保险。柯瓦斯阿默揭示了这两种态度的内在矛盾："只要心理治疗业拒绝为他们的工作提供科学证据，他们的财政压力就会加大。毕竟，如果心理治疗真的是一门艺术的话，他们应当由国家人文基金提供资助，而不是医疗保险。"与这种观点相呼应的是，考普塔、鲁格、桑德斯和霍华德（Kopta，Lueger，Saunders，& Howard，1999）在他们对心理治疗效果研究的评估中写道："某种特定疗法的效果必须通过实证方法对其有效性进行检验，才有足够的资格和说服力去向保险公司、医疗公司和政府机构申请补贴"（p. 442）。

著名的心理治疗师道·彼得森（Don Peterson）的观点也与柯瓦斯阿默的告诫不谋而合。他相当直接地指出，那些认为"虽然实证证

据都不支持，但我的经验告诉我并非如此，因此我拒绝改变我的疗法”（Peterson，1995，p.977）的治疗师“不会从我这里得到任何的同情”（p.977）。简而言之，彼得森认为心理治疗从业者必须对科学证据做出回应，否则整个行业就会被公众看做是一个“不负责任的”行业（p.977），从而丧失来自社会的支持。

本书早期版本的一些读者指出，我并没有特别强调心理学家内部的不专业行为和反科学态度在很大程度上导致了这个学科的公众形象问题，因此指责我“轻易地放过了心理学家”。好在罗宾·道斯（Robyn Dawes）在1994年出版的《纸牌搭成的屋子：建构在虚幻上的心理学和心理治疗》（*House of Cards: Psychology and Psychotherapy Built on Myth*）一书，对我持有一个平衡观点的帮助很大。谁若是对“心理学家自身就是造成此困境的重要原因”心存疑问，那就去读读这本书好了。在这部勇气可嘉的书里，道斯毫无保留地自曝家丑，并主张在专门研究人类问题的心理学里，采取科学态度对整个社会有很大的实用价值（虽然其潜力仍大有可挖）。例如，道斯写道：“确实有一门真正的心理科学，这门科学是在无数人多年以来的工作的基础上发展而来，但是，这门科学目前正因为一些从业者的行为而逐渐被忽视、贬低和遭到反对——这些从业者只是在口头上承认这门科学的存在而已”（p.vii）。

道斯等（Lilienfeld，1998；Lilienfeld，Lynn & Lohr，2003；Mook，2001；Watters & Ofshe，1999；Wood，Nezworski，Lilienfeld，& Garb，2003）所反对的是，心理学领域基于心理学的科学地位颁发资格认证，然后又用资格认证来保护心理学从业者的不科学行为。例如，一个受过良好训练的心理学家应当知道，我们有把握对总体的行为做出预测，但是在预测某个特定个人的行为时，就存在很大的不确定性（见第10章和第11章），因此，即便是最有能力的心理学家，也不应该在没有强调这点的情况下去做任何个人预测。正如道斯（1994）所言：

> 一个声称有百分之百把握预测某一个体未来行为（例如，暴力行为）的专家，注定不是一个称职的专家，因为有研究证实，不论

是一个心理健康专家，或者其他什么人，都不可能以这样的把握保证自己预测的准确性。（专业人士经常声称，尽管他们就个人来讲是接受不确定性的，但他们的专业角色“要求”他们做出这样自信的判断。不，他们不是被“要求”这样做，是他们“自愿”这样做的）。（p .vii）

简而言之，美国心理学会曾经助长了心理治疗领域的这股不正之风。这股风气让人觉得，心理学家能够通过训练获得一种“直觉洞察力”，从而能洞悉个体的行为，然而，研究证据并不支持这一观点。当有人提出质疑，认为执照制度只是一种行业限制时，该组织就把它的科学资历作为武器（一位美国心理学会主席这样回应社会人士对心理学的攻击：“我们是以科学为基础的，这就是我们有别于社会工作者、咨询师和吉普赛卜卦者的地方”；Dawes，1994，p. 21）。但是，用来维护其科学地位的这个理由却正好揭示了，有执照的心理学家具有独特的“临床洞察力”的观点是完全错误的。美国心理学会这种两面派手法催生了道斯的这本书，也在一定程度上导致了 20 世纪 80 年代美国心理协会（APS）的成立。这一协会的成员是由那些厌倦了美国心理学会“只关注蓝十字补偿费而忽视科学”的做法的心理学家所组成。

斯科特 · 利连恩费德（Scott Lilienfield），一位因其事业早期对临床心理学所做贡献而获得大卫 · 沙科夫（David Shakow）奖的学者，曾在颁奖典礼上不断重申上述观点，并警告说：“在临床心理学这一领域，我们似乎对处理伪科学这一问题完全没有兴趣，这是一个非常令人吃惊的现象，因为这个问题的火苗已经烧到我们的后院了”（Lilienfield，1998，p. 3）。他还列出了 20 世纪 90 年代在临床心理学领域泛滥成灾的几种伪科学，其中包括：

1. 用于治疗创伤的那些未经检验的怪异疗法；
2. 已经被证实是无效的、针对孤独症的一些疗法，例如辅助沟通疗法（见第 6 章）；
3. 继续使用一些未被充分验证的心理评估工具，例如各种投射测验；

4. 利用潜意识来做自我治疗的录音带；
5. 使用高度暗示性的治疗技术诱发儿时受虐的记忆。

利连恩费德援引著名临床研究者保罗·米尔（Paul Meehl）的话，“如果我们不对这一行业进行清理整顿、为我们的学生提供科学思考典范的话，外行就会替我们做”（Meehl，1993，p.728；同见 Mahrer，2000）。米尔在此指出了我们在第 11 章中讨论过的一种倾向：临床治疗师总是想让别人相信，他们拥有一些关于人类的“特殊”知识，这些知识超越了公众可获悉的、作为可重复验证的科学知识的一般行为趋势。米尔（1993）认为，临床心理学家必须更关注那些实证的公开验证的知识，并警告说，“如果认为自己拿到了博士头衔，就自以为能够在取样、感知、记录、保持、提取和推断这些人类心理受限的方面不犯错误，这是非常荒唐和自大的”（p. 728）。

心理学家约翰·佩雷兹（John Perez）认为，道斯、利连恩费德和米尔所提出的观点可以汇总成一个论点，那就是，临床心理学家应当将帮助的目标转向患者，而不是那些随心所欲地实施疗法的治疗师。正如佩雷兹（Perez，1999）所言，“我们必须决定，我们并不想创设这样一个环境：治疗师在其疗法的有效性压根没有科学依据的时候，仍然能够随心所欲地开展治疗；我们应该保障患者的权益，使其能够得到最有效的治疗”（pp. 205-206）。

然而，心理学领域依然在遭受不端行为的践踏。例如，“紧急事件应激晤谈”在许多场合被作为标准化的程序，用于治疗那些经历了爆炸、枪击、战争、恐怖主义和地震的患者（Groopman，2004；McNally, Bryant, & Ehlers，2003）。晤谈程序包括让患者“谈论事件并公开表达他们的情绪，尤其是当着也经历了同样事件的公司同事的面”（McNally et al.，2003，p. 56），其目的是为了减少创伤后应激障碍（PTSD）的发生。大多数经过晤谈的病人都报告说这种体验是有帮助的。当然，看过此书的人都不会认为其依据具有说服力（想想第 4 章中关于“安慰剂”效应的讨论）。显然，需要有一个控制组（一些没有接受紧急事件应激

晤谈的患者）。事实上，“许多创伤幸存者都在没有专业帮助的情况下从最初的创伤后反应中恢复了过来”（McNally et al.，2003，p. 45），因此需要证明，重大事件应激叙事的使用确实带来了更高的恢复率。虽然真正的控制实验所揭示的结果并非如此（Groopman，2004；McNally et al，2003），但这一疗法还在被继续使用。

艾莫瑞、奥托和奥多诺胡（Emery，Otto，& O' Donohue，2005）在搜集大量证据后所做的综述中指出，与儿童监护权相关的临床心理学中充斥着伪科学。例如，研究者描述了一些临床心理学家在儿童监护权官司中惯用的用以评估儿童最大利益的工具。在回顾了此类工具——例如，传说能够测量关系知觉和父母觉知能力的量表——之后，艾莫瑞等人（2005）做出结论：没有一个工具被证明是可靠而有效的。他们写道：“没有一个关于这些测量方法有效性的研究发表在具有同行评审机制的刊物上，而这是科学的一项重要标准”（p. 8），同时他们总结道：“我们对于这类测量最保守的评估也是尖刻的，即这些测量的构成不甚明晰，并且表现得如此糟糕，在儿童监护权评估中的运用未经任何科学的检验”（p. 7）。

艾莫瑞等人（2005）指出，不仅评估儿童监护权的工具存在缺陷，而且临床心理学家使用的概念也有问题。艾莫瑞等人举了一个所谓“双亲疏远综合症”的例子。这个概念完全基于单独个案的“临床经验”，并且缺乏科学研究结论所需要的聚合效度，但它在监护评估中却被临床心理学家当作真正的科学概念一般随心所欲地使用。

最后，认知心理学家哈尔·阿克斯（Hal Arkes）讲述了他的失落经历，联邦机构拒绝使用他所提供的关于如何使基金评审流程更加完善的方法。但是，反思之后，他不得不承认我们的专业呈现给大众的是一张窘迫的面孔。他回忆起最近收到的一个会议请柬，在这个会议上，某个分会场的专题是“与魂灵和已故祖先的沟通”，参加这一分会场的专业心理学家能够获得继续教育的资助。他承认，当专业组织也为那些可疑的“工作坊”提供资助、（并且）还有很多这类让专业蒙羞的具备职业资质的同行时，联邦机构很难接受心理学家的意见也就丝毫不奇怪了。

不过事情似乎有了一些转机。2002年一本新的杂志诞生了，它就是《心理健康实践的科学述评》(*The Scientific Review of Mental Health Practice*)(Lilienfeld，2002)。这本杂志致力于区分科学的治疗方法与那些伪科学的治疗方法，它已经得到科学心理健康实践委员会的认可。更令人振奋的是，至少有一些心理学组织已经痛下决心来整顿临床实践，并准备消除在实践过程中那种根深蒂固的“怎么样都行”的态度。利连恩费德和洛哈（Lilienfield& Lohr，2000）报告了亚利桑那州心理学资格审查委员会吊销一位心理学家执照的事件。这个心理学家试图以一种伪科学的治疗方法来治疗恐惧症，这种方法是按照预定的顺序拍打患者身体的各个部位。不用说，这种方法没有实证效度。亚利桑那州委员会命令该治疗师停止使用这种方法，并且给他“留职察看”的处罚———一个心理学组织对使用伪科学方法的成员进行查处，这样的例子在心理学界还是非常罕见的。

简言之，心理学具有像吉柯（Jekyll）和海德（Hyde）(英国作家史蒂文森的小说《化身博士》中的人物，分别代表善与恶——译者注）那样的双重人格，极端缜密的科学与伪科学及反科学的态度并存。这个学科的双重人格特性在20世纪90年代早期关于“恢复记忆－虚假记忆”的争论中表现得淋漓尽致（Brainerd & Reyna，2005；Gardner，2006；Garry，Frame，& Loftus，1999；Loftus，1997；Loftus & Guyer，2002；Loftus & Ketcham，1994；Pezdek & Banks，1996)。许多个案报告说，有患者声称回忆起几十年前当他们还是小孩的时候遭受虐待的经历，而这些记忆曾经一度被遗忘。大部分这类记忆出现在治疗干预的情境中，显然说明这些记忆中的一部分是由治疗本身所引发的（Campbell，1998；Gardner，2006；Loftus & Guyer，2002；Lynn，Loftus，Lilienfeld，& Lock，2003；Piper，1998)。有些人坚持认为这类记忆绝对不可信，另外一些人则坚称它是可信的。在这个爆炸性的社会话题所营造出的极具情绪化的氛围下，心理学家们提供了一些较为理性、平衡的意见，更为重要的是，还提供了部分关于恢复性记忆或虚假记忆的客观的实证证据（Alexander et al.，2005；Brainerd & Reyna，2005；

Bremner, Shobe, & Kihlstrom, 2000 ; Clancy, Schacter, McNally, & Pitman, 2000 ; Goodman et al., 2003 ; Mcnallly, 2003 ; Pezdek & Banks, 1996)。从这里我们能充分地看出心理学这种双重人格的特性。由治疗干预所引发的、与事实真相相反的虚假记忆中，有一部分是由某些不称职的、对科学无知的治疗师造成的，而这些治疗师都是心理学专业人士。另一方面，尽管目前对这场争论所做的结论还不够充分和确定，但这一点仍应归功于那些对相关现象实证地开展研究的心理学家的不懈努力。心理学在制造问题，同时又在解决问题！

我喜欢引用丹杰费尔德的口头禅来作为本节的题目，希望这样做能够帮我洗清"为心理学家脱罪"的恶名。心理学家道格拉斯·穆克(Douglas Mook) 在他的一本关于研究方法的书中曾提到过我借用丹杰费尔德的笑话，并且评论道："确实，通常心理学得不到应有的尊敬，但有时，它又受到了不应得的尊敬，或者因为错误的原因而受到尊敬"(Mook, 2001, P.473)。我完全同意这一感受。穆克是对的，心理学的学生应当知道这个学科所面临的窘境。就像本书中所表述的那样，作为一门研究人类行为的科学，心理学通常没有得到太多的尊敬。但是，心理学呈现给公众的印象却是很多临床治疗师宣称自己具有"独特的"洞察人心的能力——但这种洞察力在研究证据方面是站不住脚的，这一形象又使心理学获得了过多的尊敬。心理学的严谨性就在于，采取科学的方法来验证有关人类行为的各种主张；不幸的是，这一学科常由那些不尊重心理学这一严谨性的分支呈现给公众。

每个人不都是心理学家吗

我们每个人都有一套关于人类行为的理论。很难想象，如果没有这些理论，我们该怎样活下去。从这一意义上讲，我们人人都是心理学家。尽管如此，区分这种个体心理学和由心理科学所生成的知识体系仍然是十分重要的。我们将看到，这种区分之所以重要，是因为在许多大众读

物里，二者的区别经常被故意混淆了。

我们的个人心理学知识与那些对行为进行科学研究所获得的知识相比，有哪些方面的区别呢？我们已经有所讨论。我们的个人心理学知识多数是“菜谱式知识”。我们做某件事，是因为我们认为它会导致其他人做出某些相应的行为，或是因为我们相信这些事能帮助我们实现某些目标。这些都是所谓的菜谱式知识。但是，个人心理学和科学心理学（也包括一些菜谱式知识）的区别并不在于有没有菜谱式知识。最主要的区别在于，科学心理学总是力图通过实证方法检验菜谱式知识的有效性。

科学评估具有系统性和可控性，这些特性是个人评估程序所不可能具备的。事实上，心理学对于决策选择的研究表明，当行为发生的情境与原有的信念相悖时，人们就很难觉察到相关关系（见 Baron，2000）。我们只看到我们想看到的东西。心理学家已经找到出现这种现象的许多原因（Kunda，1999；Lassiter，Geers，Munhall，Ploutz-Snyder，& Breitenbecher，2002；Stanovich，2004），但是它们并非我们这里关注的重点。即使我们想在个人的基础上评估个体的菜谱式知识，那些妨碍我们对行为现象进行充分观察的先入为主的偏见，也会使我们的评估工作变得异常困难。引入科学方法的目的正是要避免个别观察者的偏见。这里的意思很简单，由科学心理学产生的菜谱式知识可能会更精确，因为和个体的菜谱式知识相比，它们经过更加严格的检验程序。

就像本章前面所讨论的那样，个体心理学和科学心理学之间的差别不仅限于对菜谱式知识的验证。科学想从自然界获得的远不止菜谱式知识。科学家们想要寻求那些能够解释药方运作机制的更为普遍的基本原理。许多人的个体心理学和科学心理学一样，也想探究更为基本的心理学规律和理论，然而这些个人理论和科学理论存在着重大的分歧。我们曾经提到过，这些个人化理论是无法证伪的。许多人的个人心理学理论缺乏缜密的建构，只是一些适用于个别情形的陈词滥调的简单堆砌，有时这些话还会自相矛盾。它们向人们保证，存在一个确定的解释，而那些与之完全对立、会彻底动摇人们信念的事件都是不可能发生的。尽管这些理论极具慰藉功能，但正如第 2 章中所讨论的，除了慰藉之外，以

这种方式提出的理论再无别的功能。这些理论都以“事后诸葛亮”的方式解释一切，对未来没有任何的预测。没有预测，也就没有给我们提供任何信息。心理学科的理论必须符合可证伪的标准，这就是心理科学与许多外行人的个人心理学的不同之处。心理学理论是能够被证伪的，因此，心理学理论蕴涵了这样一种确保其发展和进步的机制，而这是个人心理学所不具备的。

抵制科学心理学的根本原因

基于我们之前讨论过的那些理由，千万不要把个人心理学理论和科学心理学的知识混为一谈。这种混淆有时是蓄意制造出来的，目的是要诋毁心理学在公众心目中的形象。如果“人人都是心理学家”是指每一个人都有自己的心理学理论的话，那么这句话没有错。但是它常常被隐晦地暗示心理学不是一门科学。

第 1 章中已讨论过，为什么科学心理学的想法会对某些人造成威胁。一门日趋成熟的行为科学，势必会改变各类提供心理信息数据来源的个体、群体和组织。很自然，对那些长期从事人类心理和行为评论的人来说，他们肯定会抵制任何威胁其权威地位的变革。在本书的第 1 章中曾提到过，科学的进步会不断地剥夺那些原有对自然界做出解读的权威团体的地位。行星的运行、物质的本质、疾病的原因过去曾经是神学家、哲学家和通才作家把持的领域，而如今，天文学、物理学、医学、遗传学和其他学科逐渐夺取了这些主题，并将它们放置在不同的科学专门领域内。

举例来说，许多宗教都已经逐渐不再声称他们对宇宙结构具有专门的知识。除了一些局部性的争议——如特创论——科学与宗教之间的大型战争已经成为历史。科学家们探究自然世界的结构，而许多宗教则对运用这些发现时可能带来的影响做出评论，但宗教已经不再与科学争夺对于这些发现的解释权了，对有关自然界的主张的裁定权，无疑已经掌握在科学家手中。

接下来的问题就是信念评估标准的变革。不会再有新闻报纸刊登有关土星带构成的立场鲜明的社论文章。为什么呢？并没有审查机构阻止这类社论的发表。很明显，写这类社论是徒劳的。因为社会大众知道，对这一方面的知识有发言权的是科学家，而不是评论员。仅在 100 年前，报纸和那些布道坛上的牧师还曾对动物世界的物种起源学说大肆攻击。现在，这类评论大部分都消失了。科学摧毁了让任何理性思考者轻信这些观点的客观条件。心理学还将在另外一个庞大的自然领域中摧毁这类条件。100 年后，标题为“儿童早期成长，是遗传还是环境的影响？”的新闻报道可能听起来会感觉既愚蠢又过时，就像我们现在听到大主教雅舍尔（Archbishop Ussher）宣称世界是在公元前 4004 年被创造出来时的那种感觉。

有些人发现自己很难接受心理学发生的这类变革。他们顽固地坚持自己有权利对人类行为发表看法，即使这些看法与事实相去甚远。显然，“权利”用在这里并非是一个准确的措词，因为在一个自由社会里，每个人都有发表意见的权利，无论这些意见是否正确。最重要的是要意识到，许多人想要的不仅仅是发表有关人类行为见解的权利，他们真正想要的是，*无论他们说什么，人们都应该相信其所说的话*。当他们陈述一个关于人类心理学的观点时，他们希望周围的环境有利于人们接受他们的想法，这就是为什么认为心理学是“怎么着都行”的说法会有大量拥护者的原因。所谓“怎么着都行”，就是暗含心理学的主张是不能由实证方法来判别的，它只是一堆观点的集合。科学对于这种“怎么着都行”的观点来说始终是一种威胁，因为它有一系列严格的标准和程序，用以确定哪些说法是可信的。科学不是“怎么着都行”。正是这种去伪存真的能力推动了科学的进步。

简言之，许多对于科学心理学的抵制都可以归因为“*利益冲突*”。前面几章中已经讨论过，许多伪科学已经发展成为数以百万美元计的产业，它们之所以能蓬勃发展，依靠的正是公众没有意识到关于行为的主张也可以用实证方法来检验这一事实（在美国，占星师的数量是天文学家的 20 倍；见 Gilovich，1991，p. 2）。公众也没有意识到，支撑这类产

业的许多主张（如星相预测、潜意识减肥、生物节律、苦杏仁苷的使用及通灵手术），都已经被证明是无效的。美国国会下设的一个委员会曾估计过，人们每年大约花 100 亿美金在这些医疗骗术上，这让花在正规的医疗研究上的经费相形见绌（Eisenberg et al.，1993；U. S. Congress，1984）。

我们如何识别伪科学的主张？临床心理学家斯科特·利连恩费德（Lilienfeld，2005，p. 40）给出了一些注意事项，也是对本书内容的一个概括。他认为伪科学的主张有以下一些特征：

- 喜欢采用特殊的假定，使得主张免于被证伪。
- 强调主张是确证的，是不可辩驳的。
- 喜欢将提供证据的任务强加给怀疑者，而非拥护者。
- 过度依赖轶闻趣事和各类见证叙述来证实其主张。
- 逃避同行评审。
- 并非建立在已有的科学知识之上（缺乏学科关联性）。

许多伪科学术士和治疗骗术靠的就是心理学领域这种“怎么着都行”的氛围。这是一种非常容易让公众变得盲信和盲从的环境，因为，如果“怎么着都行”，公众的消费者权益就得不到保障了。正如律师彼特·哈勃所言，“在科学的边缘和科学之外……形形色色的顺势疗法药物、水晶和金字塔神奇疗效的信徒……必须借助对正统科学的诋毁来为他们的异端邪说提供立足之地”（Huber，1990，p.97）。这些兜售伪科学的人从骨子里想去掩盖这样一个事实，那就是有一套科学机制可以用于检验行为理论。迈克尔·吉瑟林（Michael Ghiselin，1989）警告说，“道理很简单，人们都试图推销特定的观点，而真正能评估观点好坏的人，不是那些在市场上推销这些观点的人”（p. 139）。在行为理论和治疗这一领域，心理学家就是那些“知道如何来评估产品”的人。这就是为什么伪心理学产业一直极力反对科学心理学在评价行为主张方面的权威性。然而，伪科学的散播者通常不与心理学家正面交锋，他们绕过心理学，带着其主张直奔媒体而去。大众传媒为那些想要绕过科学心理学的狂徒、骗子

和伪科学提供了极大的便利。泛滥的电视脱口秀节目并不要求嘉宾出示科学研究的依据。这些嘉宾只要“足够有趣”，就可以在电视上露脸。

世俗智慧通常包含许多一厢情愿的想法：人们更愿意相信世界是他们所期望的样子，而非其真实的样子。为此，科学家们承担着费力不讨好的任务，那就是去告诉公众，这个世界的本质并不是他们所想象的那样（“不，快餐对你的健康没有好处”）。媒体本来可以发挥有益的作用（告诉人们真相，而不是迎合他们的期望），然而，它们却把重心放在“娱乐”而不是提供信息上，从而使情况变得更糟。

科学确实是在把那些不符合最低检验标准的、自称是特殊知识的理论及疗法清除出局。法庭也在摒弃那些有关特殊知识的主张。在一桩著名的道伯特（Dauber ）诉梅里尔·道（Merrell Dow）药品公司的公案中，最高法院决定何时才可以在法庭上呈现专家证词——也就是说，什么才可以让专家证词具有专家性！法庭为那些考虑专家证词的法官提供了四个鉴别指标：（1）观点所基于的理论基础是“可检验的”；（2）与某一方法关联的错误率可知；（3）观点所基于的技术或方法是否经过了同行评审；（4）技术或方法是否被相关的科学团体所广泛接受（Emery, Otto, & O'Donohue, 2005）。这四个标准对应了本书的主旨：（1）可证伪性；（2）概率性预测；（3）服从同行评审的公共知识；（4）基于聚合性和共识的科学知识。法庭在排查特殊知识的主张、直觉和见证方面与科学相似。

本书曾经很简略地提及，在科学领域里的充分检验和不充分检验各是什么。内省、个人经验和见证叙述都被认为是关于人类行为主张的不充分检验。在科学心理学诞生之前，这些内容就一直被那些非心理学家的评论者视为支持其观点的宝贵证据，因此，此时会爆发冲突也就毫不奇怪了。

请不要以为我想把科学心理学刻画成一个充满敌意而让人扫兴的角色。恰恰相反，科学心理学的研究发现其实要比那些整天在媒体上反反复复、大呼小叫的伪科学有趣和精彩得多。进而，也不应该认为科学家是反对幻想和想象的，相反，在适当的情境中，科学家从来不反对幻想、

想象及浮想联翩。彼得·米德瓦（Medawar，1990）认为：

> 我坚信在适当的情境下可以夸夸其谈。我相信多数人在心理上需要保罗·詹宁斯（Paul Jennings）所说的"浮想联翩"。（你也许还记得詹宁斯在打字时，连"Bankrupt"都打成了"Bunkrapt"。任何人都需要幻想，我喜欢在听瓦格纳的歌剧或者读托尔金的小说时浮想联翩。但是，可不能将这个习惯带到科学里面（p.5）。

如果我们停下来思考一会儿，大多数人都会赞同米德瓦的观点。我们进入歌剧院或电影院时，总想看到奇幻景象，但这种情形却不太可能发生在我们去看病、买保险、到托儿所给孩子注册、坐飞机或者修理汽车的时候。这种情形也不太可能发生在我们进行心理治疗、将自己注意力缺失的孩子交给教育心理学家测试，以及把自己的朋友带到大学心理诊所进行自杀干预的时候。心理学在追求真相的过程中，必须像其他学科一样，把那些毫无依据的胡思乱想、"常识"、商业广告卖点、宗教意见、见证和妄想都清除出去。

让一门科学去告诉社会中的一部分人，他们的想法和意见是有用的，但不是在此地——这是一件相当困难的事情。心理学是科学中最后一个面临这种微妙局面的学科。这与心理学产生的时间有关。大多数学科成熟于精英控制社会结构的年代，那个时候，普通人的意见没有影响力。而心理学则产生于一个民主的传媒时代，忽视公众意见会危及自身。许多心理学家正在努力修复心理学与公众在沟通方面的糟糕记录。当越来越多的心理学家开始在与公众的沟通中发挥作用时，势必会加剧他们与那帮将个人心理学和科学心理学混为一谈人的冲突。

虽然我们每个人都有一套直觉的物理学理论，但不是每个人都是物理学家。但是，如果我们不要求让个人物理学理论取代科学物理学，就为我们每个人获悉真正的科学物理学理论（因为科学是公共性的）铺平了道路。同样道理，并非人人都是心理学家，但人人都能获得心理科学所发现的事实和理论，将之付诸实践，并丰富我们所有人对彼此的理解。

结束语

这本勾勒出“这才是心理学”的书至此已到了尾声。这一勾勒很粗略，但它对你理解心理学这门学科是如何运作的以及如何评估新的心理学主张应当有很大的帮助。我们的勾勒揭示了以下几点：

1. 心理学的进步是通过研究可解的实证问题而取得的。这种进步是不均衡的，因为心理学由许多不同的子领域构成，某些领域的问题要比其他领域具有更高的难度。
2. 心理学家提出可证伪的理论来解释他们的研究发现。
3. 理论中的概念都具有操作性定义，这些定义将随着证据的积累而逐渐演变。
4. 这些理论是通过系统实证的方法来检验的，用这种方法收集来的数据是公开的，也就是说，它允许其他科学家重复这些实验并提出批评。
5. 心理学家的数据和理论，只有在那些经过同行评审程序的科学刊物上发表之后，才算是进入了科学领域。
6. 实证主义之所以具有系统性，是因为它遵循控制和操纵的逻辑，这二者也是真实验的特性。
7. 心理学家采用许多不同的方法来获得他们的结论，这些方法的优缺点各有不同。
8. 在许多情况下，只有经过许多实验数据的缓慢积累才能得出结论。
9. 最终被揭示的行为规律，通常情况下都是一种概率关系。

1961 年，英国心理学家唐纳德·布罗德本特（Donald Broadbent）说过一段话，这段话放在今日仍然与当时一样适用。对我们来说，它可以作为如何正确看待心理学的一个精辟总结：

> 我们在此带着一份质疑暂告一段落，我们不确定将来的心理学家会持有什么信念。事情本该如此。没有人能坐在扶手椅里抓住事

实的本质，而且，在新的实验没有进行之前，我们不会知道结果如何。那些从布道坛、新闻报道以及学校的颁奖典礼听来的关于人性的言之凿凿的教条并不适合我们。相反，我们必须做好准备，去容忍目前尚不完备的关于行为的知识，但同时还要坚信客观方法的力量，终有一天，这些方法会给予我们想了解的知识。（p. 200–201）

参考文献

Adler, J. E. (1998, January). Open minds and the argument from ignorance. *Skeptical Inquirer, 22*(1), 41–44.

Alcock, J. E. (1984). Parapsychology's past eight years: A lack-of-progress report. *Skeptical Inquirer, 8*, 312–320.

Alcock, J. E. (1987). Parapsychology: Science of the anomalous or search for the soul? *Behavioral and Brain Sciences, 10*, 553–643.

Alcock, J. E. (1990). *Science and supernature: A critical appraisal of parapsychology*. Buffalo, NY: Prometheus Books.

Alexander, K. W., Quas, J. A., Goodman, G. S., Ghetti, S., Edelstein, R. S., Redlich, A. D., Cordon, I. M., & Jones, D. P. H. (2005). Traumatic impact predicts long-term memory for documented child sexual abuse. *Psychological Science, 16*, 33–40.

Alloy, L. B., Abramson, L. Y., & Francis, E. L. (1999). Do negative cognitive styles confer vulnerability to depression? *Current Directions in Psychological Science, 8*, 128–132.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Anderson, C. A., & Anderson, K. B. (1996). Violent crime rate studies in philosophical context: A destructive testing approach to heat and Southern culture of violence effects. *Journal of Personality and Social Psychology, 70*, 740–756.

Anderson, C. A., Benjamin Jr., A. J., & Bartholow, B. D. (1998). Does the gun pull the trigger? Automatic priming effects of weapon pictures and weapon names. *Psychological Science, 9*, 308–314.

Anderson, C. A., Berkowitz, L., Donnerstein, E., Huesmann, L. R., Johnson, J., Linz, D., Malamuth, N. M., & Wartella, E. (2003). The influence of media violence on youth. *Psychological Science in the Public Interest, 4*, 81–110.

Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature. *Psychological Science, 12*, 353–357.

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive

thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology, 78*, 772–790.

Anderson, C. A., & Gunderson, K. (1991). *Paul E. Meehl: Selected philosophical and methodological papers.* Minneapolis: University of Minnesota Press.

Anderson, C. A., & Huesmann, L. R. (2005). The evidence that media violence stimulates aggression in young viewers remains 'unequivocal.' *APS Observer, 18*(10), 7.

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*, 3–9.

Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., & Wright, J. C. (2001). Early childhood television viewing and adolescent behavior. *Monographs of the Society for Research in Child Development, 66*(1) (Serial No. 264), 1–158.

Angell, M., & Kassirer, J. P. (1998). Alternative medicine: The risks of untested and unregulated remedies. *The New England Journal of Medicine, 339*(12), 839–841.

Arkes, H. R. (2003). The nonuse of psychological research at two federal agencies. *Psychological Science, 14*, 5–6.

Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin, 125*, 591–600.

Asimov, I. (1989). The relativity of wrong. *Skeptical Inquirer, 14*, 35–44.

Auletta, K. (1992). *Three blind mice: How the TV networks lost their way.* New York: Vintage Books.

Austin, E. J., & Deary, I. J. (2002). Personality dispositions. In R. J. Sternberg (Ed.), *Why smart people can be so stupid* (pp. 187–211). New Haven, CT: Yale University Press.

Azar, B. (1999, November). Crowder mixes theories with humility. *APA Monitor*, p. 18.

Baron, J. (1998). *Judgment misguided: Intuition and error in public decision making.* New York: Oxford University Press.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge, MA: Cambridge University Press.

Baron-Cohen, S. (2005). *Autism: The facts.* Oxford: Oxford University Press.

Barrett, L., Dunbar, R., & Lycett, J. (2002). *Human evolutionary psychology.* Princeton, NJ: Princeton University Press.

Baumeister, R. F., Bushman, B. J., & Campbell, K. W. (2000). Self-esteem, narcissism, and aggression: Does violence result from low self-esteem or from threatened egotism? *Current Directions in Psychological Science, 9*, 26–29.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*, 1–44.

Bazerman, M. (1999). *Smart money decisions.* New York: John Wiley.

Bazerman, M. (2001). Consumer research for consumers. *Journal of Consumer Research, 27*, 499–504.

Beatty, J. (1996, May). What election '96 should be about. *The Atlantic Monthly*, pp. 114–120.

Belsky, G., & Gilovich, T. (1999). *Why smart people make big money mistakes—And how to correct them: Lessons from the new science of behavioral economics.* New York: Simon & Schuster.

Benjamin, L. T. (2001). American psychology's struggles with its curriculum:

Should a thousand flowers bloom? *American Psychologist, 56,* 735–742.
Benjamin, L. T. (2004). Behavioral science and the Nobel Prize: A history. *American Psychologist, 58,* 731–741.
Ben-Shakhar, G., Bar-Hillel, M., Blui, Y., Ben-Abba, E., & Flug, A. (1989). Can graphological analysis predict occupational success? *Journal of Applied Psychology, 71,* 645–653.
Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep. *American Psychologist, 37,* 245–257.
Berliner, D. C., & Biddle, B. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools.* Reading, MA: Addison-Wesley.
Bersoff, D. (1981). Testing and the law. *American Psychologist, 36,* 1047–1056.
Bersoff, D. (1982). Larry P. and PASE: Judicial report cards on the validity of intelligence tests. In T. R. Kratochwill (Ed.), *Advances in school psychology* (pp. 145–162). Hillsdale, NJ: Erlbaum.
Best, J. (1982). Misconceptions about psychology among students who perform highly. *Psychological Reports, 51,* 239–244.
Beyerstein, B. L. (1999). Whence cometh the myth that we use only 10% of our brains? In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 3–24). Chichester, England: John Wiley & Sons.
Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science, 10,* 399–407.
Bjorklund, D. F. (2004). *Children's thinking: Cognitive development and individual differences.* Stamford, CT: Wadsworth.
Bjorklund, D. F., & Pellegrini, A. D. (2002). *The origins of human nature: Evolutionary developmental psychology.* Washington, DC: APA.
Boden, M. A. (2003). *The creative mind: Myths and mechanisms.* London: Routledge.
Bok, S. (1974). The ethics of giving placebos. *Scientific American, 23,* 17–23.
Boneau, C. A. (1990). Psychological literacy: A first approximation. *American Psychologist, 45,* 891–900.
Bornstein, R. F. (1989). Exposure and affect. *Psychological Bulletin, 106,* 265–289.
Bower, B. (1990, July 21). The ticcing link. *Science News, 138,* 42–44.
Bower, B. (1996a, August 31). Brain images illuminate Tourette syndrome. *Science News, 150,* 133.
Bower, B. (1996b, August 24). New pitch for placebo power. *Science News, 150,* 123.
Boyer, P. J. (1999, May 17). Big guns. *The New Yorker,* pp. 54–67.
Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory.* Oxford: Oxford University Press.
Breitmeyer, B. J., & Ramey, C. T. (1986). Biological nonoptimality and quality of postnatal environment as codeterminants of intellectual development. *Child Development, 57,* 1151–1165.
Bremner, J. D., Shobe, K. K., & Kihlstrom, J. F. (2000). False memories in women with self-reported childhood sexual abuse: An empirical study. *Psychological Science, 11,* 333–337.
Broadbent, D. E. (1958). *Perception and communication.* New York: Pergamon Press.
Broadbent, D. (1961). *Behaviour.* New York: Basic Books.
Broadbent, D. (1973). *In defense of empirical psychology.* London: Methuen.
Bronfenbrenner, U., & Mahoney, M. (1975). The structure and verification of hypotheses. In U. Bronfenbrenner & M. Mahoney (Eds.), *Influence on human development* (pp. 131–152). Hinsdale, IL: Dryden.

Bronfenbrenner, U., McClelland, P., Wethington, E., Moen, P., & Ceci, S. J. (1996). *The state of Americans*. New York: Free Press.

Bronowski, J. (1956). *Science and human values*. New York: Harper & Row.

Bronowski, J. (1973). *The ascent of man*. Boston: Little, Brown.

Bronowski, J. (1974). Science, poetry, and human specificity. *American Scholar, 43*, 386–404.

Bronowski, J. (1977). *A sense of the future*. Cambridge, MA: MIT Press.

Bronowski, J. (1978a). *The common sense of science*. Cambridge, MA: Harvard University Press.

Bronowski, J. (1978b). *Magic, science, and civilization*. New York: Columbia University Press.

Budiansky, S. (1984, April 22). The meat of the matter. *Washington Post Book World*, p. 7.

Bunge, M. (1983). Speculation: Wild and sound. *New Ideas in Psychology, 1*, 3–6.

Burgess, C. A., Kirsch, I., Shane, H., Niederauer, K., Graham, S., & Bacon, A. (1998). Facilitated communication as an ideomotor response. *Psychological Science, 9*, 71–74.

Burns, B. D. (2004). Heuristics as beliefs and as behaviors: The adaptiveness of the "hot hand." *Cognitive Psychology, 48*, 295–331.

Bushman, B. J., & Anderson, C. A. (2002). Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin, 28*, 1679–1686.

Buss, D. M. (1985). Human mate selection. *American Scientist, 73*, 47–51.

Buss, D. M. (2003). *Evolutionary psychology: The new science of the mind* (2nd ed.). Boston: Allyn and Bacon.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology, 50*, 559–570.

Byrnes, G., & Kelly, I. (1992). Crisis calls and lunar cycles: A twenty-year review. *Psychological Reports, 71*, 779–795.

Byrnes, J. P. (2001). *Minds, brains, and learning: Understanding the psychological and educational relevance of neuroscientific research*. New York: Guilford Press.

Campbell, T. (1998). *Smoke and mirrors: The devastating effect of false sexual abuse claims*. New York: Plenum.

Carnoy, M., Jacobsen, R., Mishel, L., & Rothstein, R. (2005). *The charter school dust-up: Examining the evidence on enrollment and achievement*. New York: Teachers College Press.

Carpenter, S. (1999, August 21). Kansas cuts evolution from curriculum. *Science News, 156*, 117.

Carroll, M., & Nelson, T. O. (1993). Failure to obtain a generation effect during naturalistic learning. *Memory and Cognition, 21*, 361–366.

Carroll, S. B., (2005). Endless forms most beautiful: A new revolution in biology. *Skeptical Inquirer, 29*(6), 51–52.

Casscells, W., Schoenberger, A., & Graboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299*, 999–1001.

Cassidy, J. (1995, October 15). Who killed the middle class? *New Yorker*, pp. 113–124.

Catrambone, R., Jones, C. M., Jonides, J., & Seifert, C. (1995). Reasoning about curvilinear motion: Using principles or analogy. *Memory and Cognition, 23*, 368–373.

Ceci, S. J., & Hembrooke, H. (Eds.). (1998). *Expert witnesses in child abuse cases*. Washington, DC: American Psychological Association.

Chang, E. C. (1996). Cultural differences in optimism, pessimism, and coping: Predictors of subsequent adjustment in Asian American and Caucasian American college students. *Journal of Counseling Psychology, 43,* 113–123.

Chase, A. (1977). *The legacy of Malthus.* New York: Knopf.

Cheng, C. (2001). Assessing coping flexibility in real-life and laboratory settings: A multimethod approach. *Journal of Personality and Social Psychology, 80,* 814–833.

Cho, H. J., Hotopf, M., & Wessely, S. (2005). The placebo response in the treatment of chronic fatigue syndrome: A systematic review and meta-analysis. *Psychosomatic Medicine, 67,* 301–313.

Christensen, D. (1999, July 17). Does practice make perfect? The benefits of busy hospitals. *Science News, 156,* 44–45.

Christensen, D. (2001, February 3). Medicinal mimicry: Sometimes, placebos work—but how? *Science News,* pp. 74–78.

Churchland, P. M. (1988). *Matter and consciousness* (rev. ed.). Cambridge, MA: MIT Press.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul.* Cambridge, MA: MIT Press.

Clancy, S. A., Schacter, D. L., McNally, R. J., & Pitman, R. K. (2000). False recognition in women reporting recovered memories of sexual abuse. *Psychological Science, 11,* 26–31.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science.* New York: Oxford University Press.

Clotfelter, C., & Cook, P. (1989). *Selling hope: State lotteries in America.* Cambridge, MA: Harvard University Press.

Clotfelter, C., & Cook, P. (1993). The "gambler's fallacy" in lottery play. *Management Science, 39,* 1521–1525.

CNN.com (2004, May 1). Nightline airs war dead amid controversy. Retrieved December 30, 2005 from http://www.cnn.com/2004/SHOWBIZ/TV/05/01/abc.nightline/

Coan, J. A. (1997). Lost in a shopping mall: An experience with controversial research. *Ethics & Behavior, 7,* 271–284.

Cohen, R. (1985, December 17). Despair: America's double standard. *Detroit Free Press,* p. 11.

Cole, K. C. (1998). Calculated risks. *Skeptical Inquirer, 22*(5), 32–36.

Cooper, H. & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed.) (pp. 91–115). New York: Guilford Press.

Cournaud, A. (1977). The code of the scientist and its relationship to ethics. *Science, 198,* 699–705.

Cozby, P. C. (2006). *Methods in behavioral research* (9th ed.). New York: McGraw-Hill.

Crease, R. P., & Samios, N. P. (1991, January). Managing the unmanageable. *The Atlantic Monthly,* pp. 80–88.

Crews, F. (1993, November 18). The unknown Freud. *New York Review of Books,* pp. 55–66.

Crews, F. (1996). The verdict on Freud. *Psychological Science, 7,* 63–68.

Crews, F. (Ed.). (1998). *Unauthorized Freud: Doubters confront a legend.* New York: Viking.

Culver, R., Rotton, J., & Kelly, I. W. (1988). Moon mechanisms and myths. *Psychological Reports, 62,* 683–710.

Cummins, R. A., & Prior, M. P. (1992). Autism and assisted communication: A response to Bilken. *Harvard Educational Review, 62,* 228–241.

Curtis, B. M., & O'Keefe, J. H. (2002). Autonomic tone as a cardiovascular risk factor: The dangers of chronic fight or flight. *Mayo Clinic Proceedings, 77,* 45–54.

Davidow, J., & Levinson, E. M. (1993). Heuristic principles and cognitive bias in decision making: Implications for assessment in school psychology. *Psychology in the Schools, 30,* 351–361.

Davidson, J. E. & Sternberg, R. J. (Eds.). (2003). *The psychology of problem solving.* Cambridge, MA: Cambridge University Press.

Dawes, R. M. (1988). *Rational choice in an uncertain world.* San Diego: Harcourt Brace & Jovanovich.

Dawes, R. M. (1991). Probabilistic versus causal thinking. In D. Cicchetti & W. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 1, pp. 235–264). Minneapolis: University of Minnesota Press.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York: Free Press.

Dawes, R. M. (2001). *Everyday irrationality.* Boulder, CO: Westview Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1673.

Dawkins, R. (1993). Viruses of the mind. In B. Dahlbom (Ed.), *Dennett and his critics* (pp. 13–27). Cambridge, MA: Blackwell.

Dawkins, R. (1998). *Unweaving the rainbow.* Boston: Houghton Mifflin.

Day, S. X., & Rounds, J. (1998). Universality of vocational interest structure among racial minorities. *American Psychologist, 53,* 728–736.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* Oxford: Oxford University Press.

Deary, I. J. (2001). Human intelligence differences: Towards a combined experimental-differential approach. *Trends in Cognitive Sciences, 5*(4), 164–170.

deCharms, R. C., Maeda, F., Glover, G., Ludlow, D., Pauly, J., Soneji, D., Gabrieli, J., & Mackey, S. (2005). Control over brain activation and pain learned by using real-time functional MRI. *Proceedings of the National Academy of Sciences, 102,* 18626–18631.

Demetriou, A., Kui, Z. X., Spanoudis, G., Christou, C., Kyriakides, L., & Platsidou, M. (2005). The architecture, dynamics, and development of mental processing: Greek, Chinese, or universal? *Intelligence, 33,* 109–141.

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life.* New York: Simon & Schuster.

Dennett, D. C. (1998). *Brainchildren: Essays on designing minds.* Cambridge, MA: MIT Press.

Dennett, D. C. (2000). Why getting it right matters: Postmodernism and truth. *Free Inquiry, 20*(1), 40–43.

Dickson, D. H., & Kelly, I. W. (1985). The "Barnum effect" in personality assessment: A review of the literature. *Psychological Reports, 57,* 367–382.

Dillon, K. (1993, Spring). Facilitated communication, autism, and ouija. *Skeptical Inquirer, 17,* 281–287.

Distin, K. (2005). *The selfish meme.* Cambridge: Cambridge University Press.

Dixon, R., & Backman, L. (Eds.). (1995). *Compensating for psychological deficits and declines: Managing losses and promoting gain.* Mahwah, NJ: Erlbaum.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of

evolution. *American Biology Teacher, 35,* 125–129.

Dodes, J. E. (1997). The mysterious placebo. *Skeptical Inquirer, 21*(1), 44–45.

Domjan, M., & Purdy, J. E. (1995). Animal research in psychology: More than meets the eye of the general psychology student. *American Psychologist, 50,* 496–503.

Dos Passos, J. (1964). *Occasions and protests.* New York: Henry Regnery.

Dowie, J., & Elstein, A. (Eds.). (1988). *Professional judgment: A reader in clinical decision making.* New York: Cambridge University Press.

Druckman, D., & Swets, J. A. (1988). *Enhancing human performance: Issues, theories and techniques.* Washington, DC: National Academy Press.

Dudley, U. (1998). Numerology: Comes the revolution. *Skeptical Inquirer, 22*(5), 29–31.

Egan, J. (2005, December 18). Start with a stock index—Now try to turbocharge it. *New York Times,* p. BU7.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Williams, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest, 2,* 1–27.

Ehri, L. C., Nunes, S., Stahl, S., & Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's Meta-Analysis. *Review of Educational Research, 71,* 393–447.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50,* 387–395.

Eisenberg, L. (1977). The social imperatives of medical research. *Science, 198,* 1105–1110.

Eisenberg, D. M., Kessler, R., Foster, C., Norlock, F., Calkins, D., & Delbanco, T. (1993). Unconventional medicine in the United States. *The New England Journal of Medicine, 328*(4), 246–252.

Emery, R. E., Otto, R. K., & O'Donohue, W. T. (2005). A critical assessment of child custody evaluations. *Psychological Science in the Public Interest, 6*(1), 1–29.

Ernst, E., & Abbot, N. C. (1999). I shall please: The mysterious power of placebos. In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 209–213). Chichester, England: John Wiley & Sons.

Estes, W. (1979). Experimental psychology: An overview. In E. Hearst (Eds.), *The first century of experimental psychology* (pp. 12–38). Hillsdale, NJ: Erlbaum.

Evans, J. St. B. T. (2005). *How to do research: A psychologist's guide.* Hove, England: Psychology Press.

Fackelmann, K. (1996, November 9). Gastrointestinal blues. *Science News, 150,* 302–303.

Fancher, R. (1985). *The intelligence men.* New York: W. W. Norton.

Fantino, E., & Esfandiari, A. (2002). Probability matching: Encouraging optimal responding in humans. *Canadian Journal of Experimental Psychology, 56,* 58–63.

Farber, I. B., & Churchland, P. S. (1995). Consciousness and the neurosciences: Philosophical and theoretical issues. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1295–1306). Cambridge, MA: MIT Press.

Farha, B., & Steward, G. (2006). Paranormal beliefs: An analysis of college students. *Skeptical Inquirer, 30*(1), 37–40.

Faust, D., Hart, T., Guilmette, T., & Arkes, H. (1988). Neuropsychologists' capacity to detect adolescent malingerers.

Professional Psychology: Research and Practice, 19, 578–582.

Fernald, D. (1984). *The Hans legacy.* Hillsdale, NJ: Erlbaum.

Fiedler, K., & Freytag, P. (2004). Pseudocontingencies. *Journal of Personality and Social Psychology, 87,* 453–467.

Finkel, A. M. (1996, May). Who's exaggerating? *Discover, 17*(5), 48–54.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21,* 97–109.

Forrest, B., & Gross, P. R. (2004). *Creationism's Trojan Horse: The wedge of intelligent design.* Oxford: Oxford University Press.

Foster, E. A., Jobling, M. A., Taylor, P. G., Donnelly, P., Deknijff, P., Renemieremet, J., Zerjal, T., & Tyler-Smith, C. (1998). Jefferson fathered slave's last child. *Nature, 396,* 27–28.

Foster, K. R., & Huber, P. W. (1999). *Judging science: Scientific knowledge and the federal courts.* Cambridge, MA: MIT Press.

Fountain, H. (2006, January 15). On not wanting to know what hurts you. *New York Times,* p. WK14.

Fox, R. E. (1996). Charlatanism, scientism, and psychology's social contract. *American Psychologist, 51,* 777–784.

Frank, R. H. (1999). *Luxury fever: Why money fails to satisfy in an era of excess.* New York: Free Press.

Frank, R. H., & Cook, P. J. (1995). *The winner-take-all society.* New York: Free Press.

Fridson, M. S. (1993). *Investment illusions.* New York: John Wiley.

Fried, S. B. (1994). *American popular psychology: An interdisciplinary research guide.* New York: Garland Publishing.

Fried, S. B. (1998). An undergraduate course in American popular psychology. *Teaching of Psychology, 25,* 38–40.

Fried, S. B., & Schultis, G. A. (1995). *The best self-help and self awareness books.* Chicago: American Library Association.

Friedrich, J., Buday, E., & Kerr, D. (2000). Statistical training in psychology: A national survey and commentary on undergraduate programs. *Teaching of Psychology, 27,* 248–257.

Frith, U. (2003). *Autism: Explaining the enigma* (2nd ed.). Oxford: Basil Blackwell.

Gage, N. L. (1996). Confronting counsels of despair for the behavioral sciences. *Educational Researcher, 25*(3), 5–15.

Gage, N. L., & Berliner, D. (1984). *Educational psychology* (3rd ed.). San Francisco: Freeman.

Gal, I., & Baron, J. (1996). Understanding repeated simple choices. *Thinking and Reasoning, 2,* 81–98.

Galovski, T. E., Malta, L. S., & Blanchard, E. B. (2006). *Road rage: Assessment and treatment of the angry, aggressive driver.* Washington, DC: American Psychological Association.

Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory. *Psychological Science, 9,* 402–404.

Garb, H. N., & Schramke, C. J. (1996). Judgment research and neuropsychological assessment: A narrative review and meta-analysis. *Psychological Bulletin, 120,* 140–153.

Gardner, M. (1972, April). Mathematical games: Why the long arm of coincidence is usually not as long as it seems. *Scientific American, 227*(4), 110–112.

Gardner, M. (2001). Facilitated communication: A cruel farce. *Skeptical Inquirer, 25*(1), 17–19.

Gardner, M. (2006). The memory wars. *Skeptical Inquirer, 30*(1), 28–31.

Garry, M., Frame, S., & Loftus, E. F. (1999). Lie down and tell me about your childhood. In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 113–124). Chichester, England: John Wiley & Sons.

Gawande, A. (1998, February 8). No mistake. *The New Yorker*, pp. 74–81.

Gawande, A. (1999, February 9). The cancer-cluster myth. *The New Yorker*, pp. 35–37.

Gaynor, S. T. (2004). Skepticism and caricatures: B. F. Skinner turns 100. *Skeptical Inquirer, 28*(1), 26–29.

Gazzaniga, M. S. (2004). *The cognitive neurosciences* (3rd ed.). Cambridge, MA: MIT Press.

Geary, D. C. (2005). *The origin of the mind: Evolution of brain, cognition, and general intelligence.* Washington, DC: American Psychological Association.

Geschwind, N. (1985). The biology of dyslexia. In D. Gray & J. Kavanagh (Eds.), *Behavioral measures of dyslexia* (pp. 19–37). Parkton, MD: York Press.

Ghiselin, M. T. (1989). *Intellectual compromise.* New York: Paragon House.

Gigerenzer, G. (1984). External validity of laboratory experiments: The frequency-validity relationship. *American Journal of Psychology, 97,* 185–195.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you.* New York: Simon & Schuster.

Gigerenzer, G. (2004). Dread risk, September 11, and fatal traffic accidents. *Psychological Science, 15,* 286–287.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* New York: Free Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York: Cambridge University Press.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17,* 295–314.

Gladwell, M. (1996, July 8). Conquering the coma. *The New Yorker*, pp. 34–40.

Gladwell, M. (2004, December 13). The picture problem. *The New Yorker,* 74–81.

Glassner, B. (1999). *The culture of fear: Why Americans are afraid of the wrong things.* New York: Basic Books.

Gleitman, H. (1981). *Psychology.* New York: W. W. Norton.

Goldberg, L. R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender Gestalt Test. *Journal of Consulting Psychology, 23,* 25–33.

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist, 23,* 483–496.

Goldberg, L. R. (1991). Human mind versus regression equation: Five contrasts. In D. Cicchetti & W. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 1, pp. 173–184). Minneapolis: University of Minnesota Press.

Goldstein, M., & Goldstein, I. (1978). *How we know: An exploration of the scientific process.* New York: Plenum Press.

Goldstein, W. M., & Hogarth, R. M. (Eds.). (1997). *Research on judgment and decision making.* Cambridge, England: Cambridge University Press.

Goodman, G. S., Ghetti, S., Quas, J. A., Edelstein, R. S., Alexander, K. W., Redlich, A. D., Cordon, I. M., & Jones, D. P. H. (2003). A prospective study of memory for child sexual abuse:

New findings relevant to the repressed-memory controversy. *Psychological Science, 14,* 113–118.

Gosling, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin, 127,* 45–86.

Gosling, S. D., Simine, V., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59,* 93–104.

Goswami, U. (Ed.). (2004). *Blackwell handbook of childhood cognitive development.* Malden, MA: Blackwell.

Gould, S. J. (1987). Justice Scalia's misunderstanding. *Natural History, 96,* 14–21.

Greenberger, D., & Steinberg, L. (1986). *When teenagers work.* New York: Basic Books.

Greenwald, A. G., Spangenberg, E., Pratkanis, A., & Eskenazi, J. (1991). Double-blind tests of subliminal self-help audiotapes. *Psychological Science, 2,* 119–122.

Grice, G. (2001, July 30). Slice of life: How a convicted killer's corpse brought anatomy into the digital age. *The New Yorker,* pp. 36–41.

Griggs, R. A., Proctor, D. L., & Bujak-Johnson, A. (2002). The nonexistent common core. *American Psychologist, 57,* 452–453.

Groopman, J. (1999, May 10). Pet scan. *The New Yorker,* pp. 46–50.

Groopman, J. (2004, January 26). The grief industry: How much does crisis counselling help—or hurt? *The New Yorker,* pp. 30–38.

Hacking, I. (1990). *The taming of chance.* Cambridge: Cambridge University Press.

Hacohen, M. C. (2000). *Karl Popper: The formative years, 1902–1945.* Cambridge, England: Cambridge University Press.

Hafer, C. L., & Begue, L. (2005). Experimental research on just-world theory: Problems, developments, and future challenges. *Psychological Bulletin, 131,* 128–167.

Hagen, M. A. (2001). Damaged goods? What, if anything, does science tell us about the long-term effects of childhood sexual abuse? *Skeptical Inquirer, 25*(1), 54–59.

Haggbloom, S. J., Warnick, R., Warnick, J. E., Jones, V. K., Yarbrough, G. L., Russell, T. M., Borecky, C. M., McGahhey, R., Powell, J. L., Beavers, J., & Monte, E. (2002). The 100 most eminent psychologists of the 20th century. *Review of General Psychology, 6,* 139–152.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, training, and metacognitive monitoring. *American Psychologist, 53,* 449–455.

Hammond, K. R. (1996). *Human judgement and social policy.* New York: Oxford University Press.

Hammond, K. R., Harvey, L. O., & Hastie, R. (1992). Making better use of scientific knowledge: Separating truth from justice. *Psychological Science, 3,* 80–87.

Harlow, H. F. (1958). The nature of love. *American Psychologist, 13,* 673–685.

Harlow, H. F., & Suomi, S. J. (1970). The nature of love—Simplified. *American Psychologist, 25,* 161–168.

Harrington, A. (1997). *The placebo effect.* Cambridge, MA: Harvard University Press.

Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world.* Thousand Oaks, CA: Sage.

Hauser, R. M. (1998). Trends in black-white test score differentials: Uses and misuses of NAEP/SAT data. In U. Neisser (Ed.), *The rising curve:*

Long-term gains in IQ and related measures (pp. 219–249). Washington, DC: American Psychological Association.

Hearst, E. (1979). *The first century of experimental psychology.* Hillsdale, NJ: Erlbaum.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies.* Oxford: Oxford University Press.

Henriques, G. (2003). The tree of knowledge system and the theoretical unification of psychology. *Review of General Psychology, 7,* 150–182.

Henriques, G. (2004). Psychology defined. *Journal of Clinical Psychology, 60,* 1207–1221.

Henriques, G. (2005). Toward a useful mass movement. *Journal of Clinical Psychology, 61,* 121–139.

Higbee, K. L., & Clay, S. L. (1998). College students' beliefs in the ten-percent myth. *The Journal of Psychology, 132,* 469–476.

Hilton, D. J. (2003). Psychology and the financial markets: Applications to understanding and remedying irrational decision-making. In I. Brocas & J. D. Carrillo (Eds.), *The psychology of economic decisions (Vol. 1): Rationality and well-being* (pp. 273–297). Oxford: Oxford University Press.

Hines, T. M. (1998). Comprehensive review of biorhythm theory. *Psychological Reports, 83,* 19–64.

Hines, T. M. (2003). *Pseudoscience and the paranormal* (2nd ed.). Buffalo, NY: Prometheus Books.

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest, 3,* 39–77.

Holton, G. (1996). *Einstein, history, and other passions: The rebellion against science at the end of the twentieth century.* Reading, MA: Addison-Wesley.

Holton, G., & Roller, D. (1958). *Foundations of modern physical science.* Reading, MA: Addison-Wesley.

Huber, P. W. (1990). Pathological science in court. *Daedalus, 119,* 97–118.

Hudson, A., Melita, B., & Arnold, N. (1993). A case study assessing the validity of facilitated communication. *Journal of Autism and Developmental Disorders, 23,* 165–173.

Hulbert, M. (2006, January 15). Positive ratings for a ratings makeover. *New York Times,* p. BU6.

Humphrey, N. (1996). *Leaps of faith.* New York: Copernicus.

Hunt, M. (1999). *The new know-nothings: The political foes of the scientific study of human nature.* New Brunswick, NJ: Transaction Publishers.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Hyman, R. (1981). Cold reading: How to convince strangers that you know all about them. In K. Frazier (Ed.), *Paranormal borderlands of science* (pp. 79–96). Buffalo, NY: Prometheus Books.

Hyman, R. (1992). What does Goodfellow's classic analysis tell researchers today? *Journal of Experimental Psychology: General, 121,* 128–129.

Hyman, R. (1996). Evaluation of the military's twenty-year program on psychic spying. *Skeptical Inquirer, 20*(2), 21–26.

If it sounds just ducky, it probably quacks. (1990, March 23). *Detroit Free Press.*

Immen, W. (1996, August 8). Could you repeat that in Klingon? *Globe & Mail (Toronto).*

Insurance Institute for Highway Safety (2005, July 16). If you drive while phoning you're far more likely to get

into a crash in which you'll be injured. *Status Report, 40*(6), 1–3.

Ioannidis, J. (2004). Contradicted and initially stronger effects in highly cited clinical research. *JAMA, 294*, 218–228.

Jacobson, J. W., Foxx, R. M., & Mulick, J. A. (Eds.). (2004). *Controversial therapies for developmental disabilities: Fads, fashion, and science in professional practice.* Mahwah, NJ: Erlbaum.

Jacobson, J. W., Mulick, J. A., & Schwartz, A. A. (1995). A history of facilitated communication: Science, pseudoscience, and antiscience. *American Psychologist, 50*, 750–765.

Jacobson, J. W., Mulick, J. A., & Schwartz, A. A. (1996). If a tree falls in the woods. *American Psychologist, 51*, 988–989.

Jaffe, E. (2005). How random is that? Students are convenient research subjects but they're not a simple sample. *APS Observer, 18*(9), 20–30.

Johnson, H. (1991). *Sleepwalking through history: America in the Reagan years.* New York: Anchor Books.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames.* Cambridge, England: Cambridge University Press.

Kalat, J. W. (1995). *Biological psychology* (5th ed.). Pacific Grove, CA: Brooks/Cole.

Karger, H. (2005). *Shortchanged: Life and debt in the fringe economy.* San Francisco: Berrett-Koehler.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research. *American Psychologist, 56*, 405–416.

Kelly, I. W. (1997). Modern astrology: A critique. *Psychological Reports, 81*, 931–962.

Kelly, I. W. (1998). Why astrology doesn't work. *Psychological Reports, 82*, 527–546.

Kelly, I., Rotton, J., & Culver, R. (1985). The moon was full and nothing happened: A review of studies on the moon and human behavior and lunar beliefs. *Skeptical Inquirer, 10*, 129–143.

Kelly, L. (2005). *The skeptic's guide to the paranormal.* New York: Thunder's Mouth Press.

Kenrick, D. T. (2001). Evolutionary psychology, cognitive science, and dynamical systems: Building an integrative paradigm. *Current Directions in Psychological Science, 10*(1), 13–17.

Kershaw, A. (1991, July 21). The psychic guide to hiring and firing. *London Independent.*

Kety, S. (1974). Preface. In J. Swazey, *Chlorpromazine in psychiatry: A study of therapeutic intervention.* Cambridge, MA: MIT Press.

Kim, J. (1994, September). Watch out for investing newsletters luring you with

outdated returns. *Money Magazine*, pp. 12–13.

Kimble, G. A. (1999). Functional behaviorism: A plan for unity in psychology. *American Psychologist, 54*, 981–988.

King, F. (1993, March). Science literacy and public support of science. *American Psychological Society Observer, 6*(2), 2–11.

King, R. N., & Koehler, D. J. (2000). Illusory correlations in graphological inference. Journal of Experimental Psychology: *Applied, 6*, 336–348.

Kirsch, I. (Ed.). (1999). *How expectancies shape experience*. Washington, DC: American Psychological Association.

Knight, G. P., Johnson, L., Carlo, G., & Eisenberg, N. (1994). A multiplicative model of the dispositional antecedents of a prosocial behavior: Predicting more of the people more of the time. *Journal of Personality and Social Psychology, 66*, 178–183.

Koehler, J. J. (1993). The normative status of base rates at trial. In N. J. Castellan (Ed.), *Individual and group decision making* (pp. 137–149). Hillsdale, NJ: Erlbaum.

Kopta, S. M., Lueger, R. J., Saunders, S. M., & Howard, K. I. (1999). Individual psychotherapy outcome and process research. *Annual Review of Psychology, 50*, 441–469.

Kramer, S. P. (1987). *How to think like a scientist: Answering questions by the scientific method*. New York: Thomas Y. Crowell Junior Books.

Krauthammer, C. (1985, December 29). The twilight of psychotherapy? *Detroit News*.

Krippner, S. (1977). *Advances in parapsychological research: 1. Psychokinesis*. New York: Plenum Press.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation. in a juror reasoning task. *Psychological Science, 5*, 289–296.

Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.

Kushner, H. I. (1999). *A cursing brain? The histories of Tourette Syndrome*. Cambridge, MA: Harvard University Press.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32*, 311–328.

Langewiesche, W. (1993, December). The turn. *The Atlantic Monthly*, pp. 115–122.

Lassiter, G. D., Geers, A. L., Munhall, P. J., Ploutz-Snyder, R. J., & Breitenbecher, D. L. (2002). Illusory causation: Why it occurs. *Psychological Science, 13*, 299–305.

Latane, B., & Darley, J. (1970). *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts.

Lazar, I., Darlington, R., Murray, H., Royce, J., & Sniper, A. (1982). Lasting effects of early education: A report from the Consortium of Longitudinal Studies. *Monographs of the Society for Research in Child Development, 47* (Serial No. 195).

Leary, D. E. (2001). One big idea, one ultimate concern. *American Psychologist, 56*, 425–432.

Leavy, J. (1992). Our spooky presidential coincidences contest. *Skeptical Inquirer, 16*, 316–319.

Lee, V., Brooks-Gunn, J., Schnur, E., & Liaw, F. (1990). Are Head Start's effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development, 61*, 495–507.

Lees-Haley, P. R. (1997). Manipulation of perception in mass tort litigation.

Natural Resources & Environment, 12, 64–68.

Leibowitz, H. W. (1996). The symbiosis between basic and applied research. *American Psychologist, 51,* 366–370.

Lerner, L. S. (2005). What should we think about Americans' beliefs regarding evolution? *Skeptical Inquirer, 29*(6), 59–60.

Leslie, A. M. (2000). "Theory of mind" as a mechanism of selective attention. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 1235–1247). Cambridge, MA: MIT Press.

Levin, J. R., & O'Donnell, A. M. (2000). What to do about educational research's credibility gaps? *Issues in Education: Contributions from Educational Psychology, 5,* 1–87.

Levy, S. (2005, January 31). Does your iPod play favorites? *Newsweek,* p. 10.

Lewis, M. (1997). *Trail fever.* New York: Knopf.

Li, C. (1975). *Path analysis: A primer.* Pacific Grove, CA: Boxwood Press.

Lilienfeld, S. O. (1998). Pseudoscience in contemporary clinical psychology: What it is and what we can do about it. *The Clinical Psychologist, 51*(4), 3–9.

Lilienfeld, S. O. (1999). Projective measures of personality and psychopathology: How well do they work? *Skeptical Inquirer, 23*(5), 32–39.

Lilienfeld, S. O. (2002). New journal dedicated to distinguishing science from pseudoscience in mental health. *Skeptical Inquirer, 26*(3), 7–8.

Lilienfeld, S. O. (2005). The 10 commandments of helping students distinguish science from pseudoscience in psychology. *APS Observer, 18*(9), 39–51.

Lilienfeld, S. O., & Lohr, J. M. (2000). Thought field therapy practitioners and educators sanctioned. *Skeptical Inquirer, 24*(2), 5.

Lilienfeld, S. O., Lohr, J. M., & Moirer, D. (2001). The teaching of courses in the science and pseudoscience of psychology: Useful resources. *Teaching of Psychology, 28*(3), 182–191.

Lilienfeld, S. O., Lynn, S. J., & Lohr, J. M. (Eds.). (2003). *Science and pseudoscience in clinical psychology.* New York: Guilford Press.

Lilienfeld, S. O.,Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1,* 27–66.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48,* 1181–1209.

Loftus, E. F. (1997, September). Creating false memories. *Scientific American, 276*(9), 70–75.

Loftus, E. F., & Guyer, M. J. (2002, May/June). Who abused Jane Doe: The hazards of the single case history. *Skeptical Inquirer, 26*(3), 24–32.

Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memory: False memories and allegations of sexual abuse.* New York: St. Martin's Press.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "General Intelligence, Objectively Determined and Measured." *Journal of Personality and Social Psychology, 86,* 96–111.

Lynn, S. J., Loftus, E. F., Lilienfeld, S. O., & Lock, T. (2003). Memory recovery techniques in psychotherapy: Problems and pitfalls. *Skeptical Inquirer, 27*(4), 40–46.

MacCoun, R. (2002, December). Why a psychologist won the Nobel Prize in economics. *APS Observer, 15*(10), 1–8.

MacDonald, A. (1990, October 3). Parents fear wrong things, survey suggests.

Ann Arbor News (taken from The New York Times wire service).

Machado, A., Lourenço, O., & Silva, F. J. (2000). Facts, concepts, and theories: The shape of psychology's epistemic triangle. *Behavior and Philosophy, 28,* 1–40.

Macmillan, M. (1997). *Freud evaluated.* Cambridge, MA: MIT Press.

Madrick, J. (2006, January 12). The way to a fair deal. *New York Review of Books,* pp. 37–40.

Magee, B. (1985). *Philosophy and the real world: An introduction to Karl Popper.* LaSalle, IL: Open Court.

Mahrer, A. R. (2000). Philosophy of science and the foundations of psychotherapy. *American Psychologist, 55,* 1117–1125.

Maizels, M. (2005). Why should physicians care about behavioral research? *Headache, 45,* 411–413.

Malkiel, B. G. (2004). *A random walk down Wall Street* (8th ed.). New York: Norton.

Margolis, H. (1996). *Dealing with risk.* Chicago: University of Chicago Press.

Marks, D., & Kammann, R. (1980). *The psychology of the psychic.* Buffalo, NY: Prometheus Books.

Marks, D. F. (2001). *The psychology of the psychic.* Buffalo, NY: Prometheus Books.

Marshall, E. (1980). Police science and psychics. *Science, 210,* 994–995.

Martin, B. (1998). Coincidences: Remarkable or random? *Skeptical Inquirer, 22*(5), 23–28.

Matthews, K. A. (2005). Psychological perspectives on the development of coronary heart disease. *American Psychologist, 60,* 783–796.

Maynard Smith, J. (1998). *Evolutionary genetics* (2nd ed.). Oxford: Oxford University Press.

McBride-Chang, C., & Kail, R. V. (2002). Cross-cultural similarities in the predictors of reading acquisition. *Child Development, 73,* 1392–1407.

McCloskey, M. (1983, April). Intuitive physics. *Scientific American, 248*(4), 122–130.

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence.* Wellesley, MA: AK Peters.

McCullough, M. L. (2001). Freud's seduction theory and its rehabilitation: A saga of one mistake after another. *Review of General Psychology, 5,* 3–22.

McCutcheon, L. E., Furnham, A., & Davis, G. (1993). A cross-national comparison of students' misconceptions about psychology. *Psychological Reports, 72,* 243–247.

McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodword, M., Haworth, C., Palamara, P., & Cercarelli, R. (2005, August 20). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. *British Medical Journal, 331*(7514), 428.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50,* 215–241.

McGraw, K. O.,Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science, 11,* 502–506.

McNally, R. J. (2003). Recovering memories of trauma: A view from the laboratory. *Current Directions in Psychological Science, 12,* 32–35.

McNally, R. J., Bryant, R. A., & Ehlers, A. (2003). Does early psychological intervention promote recovery from posttraumatic stress? *Psychological Science in the Public Interest, 4*(2), 45–79.

Medawar, P. B. (1967). *The art of the soluble.* London: Methuen.

Medawar, P. B. (1979). *Advice to a young scientist.* New York: Harper & Row.

Medawar, P. B. (1984). *The limits of science.* New York: Harper & Row.

Medawar, P. B. (1990). *The threat and the glory.* New York: HarperCollins.

Medawar, P. B., & Medawar, J. S. (1983). *Aristotle to zoos: A philosophical dictionary of biology.* Cambridge, MA: Harvard University Press.

Medin, D. L., & Atran, S. (2004). The native mind: Biological categorization and reasoning in development and across cultures. *Psychological Review, 111,* 960–983.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the literature.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50,* 370–375.

Meehl, P. E. (1993). Philosophy of science: Help or hindrance? *Psychological Reports, 72,* 707–733.

Metalsky, G. I., & Joiner, T. E. (1992). Vulnerability to depressive symptomatology: A prospective test of the diathesis-stress and causal mediation components of the hopelessness theory of depression. *Journal of Personality and Social Psychology, 63,* 667–675.

Miller, M., Azrael, D., & Hemenway, D. (2002). Firearm availability and unintentional firearm deaths, suicide, and homicide among 5–14 year olds. *Journal of Trauma, Injury, Infection, and Critical Care, 52,* 267–275.

Miller, N. E. (1985a). Rx: Biofeedback. *Psychology Today, 19,* 54–59.

Miller, N. E. (1985b). The value of behavioral research on animals. *American Psychologist, 40,* 423–440.

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin, 125,* 387–391.

Mishel, L. (1995, Fall). Rising tides, sinking wages. *American Prospect,* pp. 60–64.

Monk, R. (1990). *Ludwig Wittgenstein: The duty of genius.* New York: Free Press.

Mook, D. G. (1982). *Psychological research: Strategies and tactics.* New York: Harper & Row.

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38,* 379–387.

Mook, D. G. (1989). The myth of external validity. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 25–43). Cambridge, England: Cambridge University Press.

Mook, D. G. (2001). *Psychological research: The ideas behind the methods.* New York: Norton.

Moore, T. E. (1995). Subliminal self-help auditory tapes: An empirical test of perceptual consequences. *Canadian Journal of Behavioural Science, 27,* 9–20.

Moore, T. E. (1996). Scientific consensus and expert testimony: Lessons from the Judas Priest trial. *Skeptical Inquirer, 20*(6), 32–38.

Mostert, M. P. (2001). Facilitated communication since 1995: A review of published studies. *Journal of Autism and Developmental Disorders, 31*(3), 287–313.

Mulick, J., Jacobson, J., & Kobe, F. (1993, Spring). Anguished silence and helping hands: Autism and facilitated communication. *Skeptical Inquirer, 17,* 270–280.

Murchison, C. (1934). *Handbook of general experimental psychology.* Worcester, MA: Clark University Press.

Musella, D. P. (2005). Gallup Poll shows that Americans' belief in the paranormal persists. *Skeptical Inquirer, 29*(5), 5.

Nathan, P. E., & Gorman, J. M. (Ed.). (1998). *A guide to treatments that work.* New York: Oxford University Press.

National Highway Traffic Safety Administration (2004). *Traffic safety facts, 2004.* Washington, DC: Author.

National Reading Panel: Reports of the Subgroups. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC.

National Safety Council. (1990). *Accident facts* (1990 edition). Chicago: Author.

National Safety Council (2001). *Report on Injuries in America, 2001.* Retrieved March 29, 2002, from www.nsc.org/library/rept2000.htm

Neter, E., & Ben-Shakhar, G. (1989). Predictive validity of graphological inferences: A meta-analysis. *Personality and Individual Differences, 10,* 737–745.

Nickerson, R. (1992). *Looking ahead: Human factors challenges in a changing world.* Hillsdale, NJ: Erlbaum.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220.

Nickerson, R. S. (1999). Basic versus applied research. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 409–423). Cambridge, MA: MIT Press.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review, 109,* 330–357.

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning.* Mahwah, NJ: Erlbaum.

Niemeyer, W., & Starlinger, I. (1981). Do the blind hear better? *Audiology, 20,* 503–515.

Nisbett, R. E. (2003). *The geography of thought.* New York: Free Press.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111,* 455–485.

Olson, R. K. (2002). Dyslexia: Nature and nurture. *Dyslexia, 8,* 143–159.

Olson, R. K. (2004). SSSR, environment, and genes. *Scientific Studies of Reading, 8,* 111–124.

Orkin, M. (2000). *What are the odds? Chance in eveyday life.* New York: W. H. Freeman.

Page, E., & Keith, T. (1981). Effects of U.S. private schools: A technical analysis of two recent claims. *Educational Researcher, 10,* 7–17.

Paik, H., & Comstock, G. (1994). The effects of television violence on antisocial behavior: A meta-analysis. *Communications Research, 21,* 516–546.

Paloutzian, R. F. (1983). *Invitation to the psychology of religion.* Glenview, IL: Scott, Foresman.

Paul, A. M. (2001, March/April 2001). Self-help: Shattering the myths. *Psychology Today,* pp. 62–68.

Paul, G. L. (1966). *Insight vs. desensitization in psychotherapy.* Stanford, CA: Stanford University Press.

Paul, G. L. (1967). Insight vs. desensitization in psychotherapy two years after termination. *Journal of Consulting Psychology, 31,* 333–348.

Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences.* New York: Vintage Books.

Pennington, B. F., & Olson, R. K. (2005). Genetics of dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 453–472). Malden, MA: Blackwell.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental

psychopathology. *Journal of Child Psychology and Psychiatry, 37,* 51–87.

Perez, J. E. (1999). Clients deserve empirically supported treatments, not romanticism. *American Psychologist, 54,* 205–206.

Peterson, D. R. (1995). The reflective educator. *American Psychologist, 50,* 975–983.

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research, 28,* 450–461.

Petry, N. M. (2005). *Pathological gambling: Etiology, comorbidity, and treatment.* Washington, DC: American Psychological Association.

Pettit, G. S., Bates, J. E., Dodge, K. A., & Meece, D. W. (1999). The impact of after-school peer contact on early adolescent externalizing problems is moderated by parental monitoring, perceived neighborhood safety, and prior adjustment. *Child Development, 70,* 768–778.

Pezdek, K., & Banks, W. P. (Eds.). (1996). *The recovered memory/false memory debate.* San Diego: Academic Press.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence.* Cambridge, MA: MIT Press.

Pigliucci, M. (2002). Hypothesis testing and the nature of skeptical investigations. *Skeptical Inquirer, 26*(6), 27–48.

Pinker, S. (1997). *How the mind works.* New York: W. W. Norton.

Pinker, S. (2002). *The blank slate: The modern denial of human nature.* New York: Viking.

Piper, A. (1998). Multiple personality disorder: Witchcraft survives in the twentieth century. *Skeptical Inquirer,* 22(3), 44–50.

Plotkin, D. (1996, June). Good news and bad news about breast cancer. *Atlantic Monthly,* 53–82.

Popper, K. R. (1959). *The logic of scientific discovery.* New York: Harper & Row.

Popper, K. R. (1963). *Conjectures and refutations.* New York: Harper & Row.

Popper, K. R. (1972). *Objective knowledge.* Oxford, England: Oxford University Press.

Popper, K. R. (1976). *Unended quest: An intellectual biography.* La Salle, IL: Open Court.

Postman, N. (1988). *Conscientious Objections.* New York: Vintage Books.

Powell, B. (1993, December). Sloppy reasoning, misused data. *Phi Delta Kappan, 75*(4), 283, 352.

Powell, B., & Steelman, L. C. (1996). Bewitched, bothered, and bewildering: The use and misuse of state SAT and ACT scores. *Harvard Educational Review, 66,* 27–59.

Pressley, M. (2005). *Reading instruction that works: The case for balanced teaching* (3rd ed.). New York: Guilford Press.

Price, C. J., & McCrory, E. (2005). Functional brain imaging studies of skilled reading and developmental dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 473–496). Malden, MA: Blackwell.

Proctor, R. W., & Capaldi, E. J. (2001). Empirical evaluation and justification of methodologies. *Psychological Bulletin, 127,* 759–772.

Radford, B. (1999). The ten-percent myth. *Skeptical Inquirer, 23*(2), 52–53.

Radford, B. (2005). Ringing false alarms: Skepticism and media scares. *Skeptical Inquirer, 29*(2), 34–39.

Radford, B. (2006). Geller revisited. *Skeptical Inquirer, 30*(1), 27.

Ramey, S. L. (1999). Head Start and preschool education. *American Psychologist, 54,* 344–346.

Randall, L. (2005, September 18). Dangling particles. *New York Times,* p. WK13.

Randi, J. (1983). The project alpha experiment: 1. The first two years. *Skeptical Inquirer, 7*, 24–33.

Randi, J. (1995). *An encyclopedia of claims, frauds, and hoaxes of the occult and supernatural exposed by James Randi.* New York: St. Martin's Press.

Randi, J. (2005). Fakers and innocents: The one million dollar challenge and those who try for it. *Skeptical Inquirer, 29*(4), 45–50.

Raymo, C. (1999). *Skeptics and true believers.* Toronto: Doubleday Canada.

Rayner, K. (1998). Eye movements in reading and information processing: 20 Years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*, 31–74.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2002, March). How should reading be taught? *Scientific American, 286*(3), 84–91.

Redding, R. E. (1998). How commonsense psychology can inform law and psycholegal research. *University of Chicago Law School Roundtable, 5*, 107–142.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine, 336*(7), 453–458.

Redelmeier, D. A., & Tibshirani, R. J. (2001). Car phones and car crashes: Some popular misconceptions. *Canadian Medical Association Journal, 164*(11), 1581–1582.

Rice, T. W. (2003). Believe it or not: Religious and other paranormal beliefs in the United States. *Journal for the Scientific Study of Religion, 42*, 95–106.

Ridley, M. (1996). *Evolution* (2nd ed.). Cambridge, MA: Blackwell Science.

Ridley, M. (1999). *Genome: The autobiography of a species in 23 chapters.* New York: HarperCollins.

Riener, C., Proffitt, D. R., & Salthouse, T. (2005). A psychometric approach to intuitive physics. *Psychonomic Bulletin and Review, 12*, 740–745.

Rind, B., Tromovitch, P., & Bauserman, R. (2001). The validity and appropriateness of methods, analyses, and conclusions in Rind et al. (1998): A rebuttal of victimological critique from Ondersma et al. (2001) and Dallam et al. (2001). *Psychological Bulletin, 127*, 734–758.

Robins, R. W., Gosling, S. D., & Craik, K. H. (1999). An empirical analysis of trends in psychology. *American Psychologist, 54*, 117–128.

Robins, R. W., Gosling, S. D., & Craik, K. H. (2000). Trends in psychology: An empirical issue. *American Psychologist, 55*, 276–277.

Robinson, D. N. (2001). Sigmund Koch—Philosophically speaking. *American Psychologist, 56*, 420–424.

Robinson, W. S. (1992). *Computers, minds, and robots.* Philadelphia: Temple University Press.

Rosa, L., Rosa, E., Sarner, L., & Barrett, S. (1998). A close look at therapeutic touch. *JAMA, 279*, 1005–1010.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 46*, 775–776.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*, 183–192.

Ross, J. F. (1999). *The polar bear strategy.* New York: Perseus Books.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* Philadelphia: Temple University Press.

Rotton, J., & Kelly, I. W. (1985). Much ado about the full moon. *Psychological Bulletin, 97,* 286–306.

Rowe, W. (1993). Psychic detectives: A critical examination. *Skeptical Inquirer, 17,* 159–165.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD Triad Hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76,* 574–586.

Ruscio, J. (2000). Risky business: Vividness, availability, and the media paradox. *Skeptical Inquirer, 24*(2), 22–26.

Ruscio, J. (2002, March). The emptiness of holism. *Skeptical Inquirer, 26*(2), 46–50.

Ruse, M. (1999). *Mystery of mysteries: Is evolution a social construction?* Cambridge, MA: Harvard University Press.

Russo, E. (2002, December 9). The biological basis of the placebo effect. *The Scientist, 16*(24), 30–31.

Russo, F. (1999, May). The clinical-trials bottleneck. *The Atlantic Monthly,* pp. 30–36.

Rutter, M. (1979). Maternal deprivation, 1972–1978: New findings, new concepts, and new approaches. *Child Development, 50,* 283–305.

Rzewnicki, R., & Forgays, D. G. (1987). Recidivism and self-cure of smoking and obesity: An attempt to replicate. *American Psychologist, 42,* 97–100.

Sagan, C. (1996). *The demon-haunted world: Science as a candle in the dark.* New York: Random House.

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science, 13,* 140–144.

Samuels, D. (2004, October 3). Congress 101: If you want success, don't mess with the gun lobby. *New York Times,* p. 10.

Sandoval, J., & Miille, M. (1980). Accuracy of judgments of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology, 48,* 249–253.

Santrock, J. W., Minnett, A. M., & Campbell, B. D. (1994). *The authoritative guide to self-help books.* New York: Guilford Press.

Schachter, S. (1982). Recidivism and self-cure of smoking and obesity. *American Psychologist, 37,* 436–444.

Scott, E. C. (2005). *Evolution vs. creationism.* Berkeley, CA: University of California Press.

Scott, S. (1999, January 2). Risking all on alternative cancer therapies. *National Post (Toronto),* p. B1.

Seligman, M. E. P. (2002). *Authentic happiness.* New York: Free Press.

Senechal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73,* 445–460.

Shadish, W. R., & Baldwin, S. A. (2005). Effects of behavioral marital therapy: A meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology, 73,* 6–14.

Shaffer, L. (1981). The growth and limits of recipe knowledge. *Journal of Mind and Behavior, 2,* 71–83.

Shane, H. (1993, June). FC: Facilitated or "factitious" communication. *Communicating Together, 11*(2), 11–13.

Shapiro, A. (1960). A contribution to a history of the placebo effect. *Behavioral Science, 5,* 109–135.

Shapiro, A., Shapiro, E., Bruun, R., & Sweet, R. (1978). *Gilles de la Tourette syndrome.* New York: Raven Press.

Shaywitz, S. E., & Shaywitz, B. A. (2004). Neurobiologic basis for reading and reading disability. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence*

in reading research (pp. 417–442). Baltimore: Paul Brookes.

Shepard, R. (1983). "Idealized" figures in textbooks versus psychology as an empirical science. *American Psychologist, 38,* 855.

Shermer, M. (1997). *Why people believe weird things.* New York: W. H. Freeman.

Shermer, M. (2005). *Science friction: Where the known meets the unknown.* New York: Times Books.

Siegert, R. J., & Ward, T. (2002). Clinical psychology and evolutionary psychology: Toward a dialogue. *Review of General Psychology, 6,* 235–259.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development, 64*(Serial No. 256), 1–114.

Simmons, R., Burgeson, R., Carlton-Ford, S., & Blyth, D. (1987). The impact of cumulative change in early adolescence. *Child Development, 58,* 1220–1234.

Simonton, D. K. (2004). Psychology's status as a scientific discipline: Its empirical placement within an implicit hierarchy of the sciences. *Review of General Psychology, 8,* 59–67.

Sinaceur, M., Heath, C., & Cole, S. (2005). Emotional and deliberative reactions to a public crisis: Mad Cow Disease in France. *Psychological Science, 16,* 247–254.

Sivak, M., & Flannagan, M. J. (2003). Flying and driving after the September 11 attacks. *American Scientist, 91,* 6–7.

Slemrod, J., & Bakija, J. (1996). *Taxing ourselves: A citizen's guide to the great debate over tax reform.* Cambridge, MA: MIT Press.

Smith, D. (2003). Angry thoughts, at-risk hearts. *APA Monitor, 34*(3), 46–48.

Smith, M., & Belcher, R. (1993). Facilitated communication with adults with autism. *Journal of Autism and Developmental Disorders, 23,* 175–183.

Smith, R. E., Wheeler, G., & Diener, E. (1975). Faith without works. *Journal of Applied Social Psychology, 5,* 320–330.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Snowling, M. J., & Hulme, C. (Eds.). (2005). *The science of reading: A handbook.* Malden, MA: Blackwell.

Spanos, N. P. (1996). *Multiple identities and false memories: A sociocognitive perspective.* Washington, DC: American Psychological Association.

Spitz, H. H. (1997). *Nonconscious movements: From mystical messages to facilitated communication.* Mahwah, NJ: Erlbaum.

Stankov, L., & Spilsbury, G. (1978). The measurement of auditory abilities of blind, partially sighted, and sighted children. *Applied Psychological Measurement, 2,* 491–503.

Stanovich, K. E. (1989). Implicit philosophies of mind: The dualism scale and its relationships with religiosity and belief in extrasensory perception. *Journal of Psychology, 123,* 5–23.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.

Stanovich, K. E. (2000). *Progress in understanding reading.* New York: Guilford Press.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin.* Chicago: University of Chicago Press.

Stanovich, K. E., & West, R. F. (1998). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory & Cognition, 28,* 161–179.

Steinberg, L., Brown, B. B., & Dornbusch, S. M. (1996). *Beyond the classroom: Why school reform has failed and what parents need to do*. New York: Simon & Schuster.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (Ed.). (2002). *Why smart people can be so stupid*. New Haven, CT: Yale University Press.

Sternberg, R. J. (Ed.). (2005). *Unity in psychology: Possibility or pipedream?* Washington, DC: American Psychological Association.

Sternberg, R. J., & Grigorenko, E. L. (Eds.). (2002). *The general factor of intelligence: How general is it?* Mahwah, NJ: Lawrence Erlbaum Associates.

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49*, 479–502.

Stevenson, H., Stigler, J., Lee, S., Lucker, G., Kitamura, S., & Hsu, C. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development, 56*, 718–734.

Stewart-Williams, S., & Podd, J. (2004). The placebo effect: Dissolving the expectancy versus conditioning debate. *Psychological Bulletin, 130*, 324–340.

Stout, M. (2000). *The feel-good curriculum: The dumbing down of America's kids in the name of self-esteem*. Cambridge, MA: Perseus Publishing.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and and conversing on a cellular telephone. *Pschological Science, 12*, 462–466.

Suls, J., & Bunde, J. (2005). Anger, anxiety, and depression as risk factors for cardiovascular disease: The problems and implications of overlapping affective dispositions. *Psychological Bulletin, 131*, 260–300.

Sunstein, C. R. (2002). *Risk and reason: Safety, law, and the environment*. Cambridge: Cambridge University Press.

Sutherland, S. (1992). *Irrationality: The enemy within*. London: Constable.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*, 143–148.

Swanson, D. (2001). *Nibbling on Einstein's brain: The good, the bad & the bogus in science*. Toronto: Annick Press.

Swanson, D. (2004). *Turn it loose: The scientist in absolutely everybody*. Toronto: Annick Press.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1–26.

Taleb, N. N. (2001). *Fooled by randomness*. New York: Texere.

Talbot, M. (2005, December 5). Darwin in the dock. *The New Yorker*, pp. 66–77.

Tassoni, C. J. (1996). Representativeness in the market for bets on National Football League games. *Journal of Behavioral Decision Making, 9*, 115–124.

Taube, K. T., & Linden, K. W. (1989). State mean SAT score as a function of participation rate and other educational and demographic variables. *Applied Measurement in Education*, 2, 143–159.

Tetlock, P. E. (2005). *Expert political judgment*. Princeton, NJ: Princeton University Press.

Thaler, R. H. (1992). *The winner's curse: Paradoxes and anomalies of economic life*. New York: Free Press.

Thomas, G., Alexander, K., & Eckland, B. (1979). Access to higher education: The importance of race, sex, social class, and academic credentials. *School Review, 87*, 133–156.

Thornton, E. (1986). *The Freudian fallacy*. London: Paladin Books.

Tooby, J. (2002, October 6). The greatest Englishman since Newton. *New York Times Book Review*, p. 12.

Toplak, M., Liu, E., Macpherson, R., Toneatto, T., & Stanovich, K. E. (in press). The reasoning skills and thinking dispositions of problem gamblers: A dual-process taxonomy. *Journal of Behavioral Decision Making.*

Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments & Computers, 30,* 583–591.

Tufte, E. R. (1997). *Visual and statistical thinking: Displays of evidence for making decisions.* Cheshire, CT: Graphics Press.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59,* 433–460.

Turner, C., Simons, L., Berkowitz, L., & Frodi, A. (1977). The stimulating and inhibiting effects of weapons on aggressive behavior. *Aggressive Behavior, 3,* 355–378.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Twachtman-Cullen, D. (1997). *A passion to believe.* Boulder, CO: Westview.

U.S. Congress House Select Committee on Aging. (1984, May 31). *Quackery: A $10 billion scandal.* Washington, DC: U.S. Government Printing Office.

University of California, Berkeley. (1991, January). The 18-year gap. *Berkeley Wellness Letter,* p. 2.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33,* 67–81.

Updegrave, W. L. (1995, August). Why funds don't do better. *Money Magazine,* pp. 58–65.

Vaughan, E. D. (1977). Misconceptions about psychology among introductory psychology students. *Teaching of Psychology, 4,* 138–141.

Vazire, S., & Gosling, S. D. (2003). Bridging psychology and biology with animal research. *American Psychologist, 58,* 407–408.

Vellutino, F., Fletcher, J. M., Snowling, M., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45,* 2–40.

Violato, C. (1984). Effects of Canadianization of American-biased items on the WAIS and WAIS-R information subtests. *Canadian Journal of Behavioral Science, 16,* 36–41.

Violato, C. (1986). Canadian versions of the information subtests of the Wechsler tests of intelligence. *Canadian Psychology, 27,* 69–74.

Wachs, T. D. (2000). *Necessary but not sufficient: The respective roles of single and multiple influences on individual development.* Washington, DC: American Psychological Association.

Wagenaar, W. A. (1988). *Paradoxes of gambling behavior.* Hove, England: Erlbaum.

Wagenaar, W. A., & Keren, G. (1986). The seat belt paradox: Effect of adopted roles on information seeking. *Organizational Behavior and Human Decision Processes, 38,* 1–6.

Wainer, H. (1989). Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics, 14,* 121–140.

Wainer, H. (1993). Does spending money on education help? A reaction to the Heritage Foundation and the Wall Street Journal. *Educational Researcher, 22*(9), 22–24.

Wainer, H. (1999). The most dangerous profession: A note on nonsampling error. *Psychological Methods, 4,* 250–256.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin, 122,* 203–215.

Watters, E., & Ofshe, R. (1999). *Therapy's delusions: The myth of the unconscious and the exploitation of the walking worried.* New York: Scribners.

Wegner, D. M., Fuller, V. A., & Sparrow, B. (2003). Clever hands: Uncontrolled intelligence in facilitated communication. *Journal of Personality and Social Psychology, 85,* 5–19.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684.

Wheeler, D. L., Jacobson, J. W., Paglieri, R. A., & Schwartz, A. A. (1993). An experimental assessment of facilitated communication. *Mental Retardation, 31,* 49–60.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 595–604.

Wilson, E. O. (1998). *Consilience: The unity of knowledge.* New York: Knopf.

Wilson, R. A., & Keil, F. C. (Eds.). (1999). *The MIT Encyclopedia of the cognitive sciences.* Cambridge, MA: MIT Press.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

Wintre, M. G., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology, 42,* 216–225.

Wood, J. M., Nezworski, M. T., Lilienfeld, S. O., & Garb, H. N. (2003). *What's wrong with the Rorschach?* San Francisco: Jossey-Bass.

Wood, J. M., Nezworski, T., & Stejskal, W. J. (1996). The comprehensive system for the Rorschach: A critical examination. *Psychological Science, 7,* 3–10.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests, Revised-Normative Update.* Shoreview, MN: American Guidance Service.

Woolf, A., & Lesperance, L. (2003, September 22). What should we worry about? *Newsweek,* p. 72.

Woolley, B. (2000). *The bride of science: Romance, reason, and Byron's daughter.* New York: McGraw-Hill.

Wright, R. (1988). *Three scientists and their gods.* New York: Harper & Row.

Wright, R. (1999, December 13). The accidental creationist. *The New Yorker,* pp. 56–65.

Wrightsman, L. S. (1999). *Judicial decision making: Is psychology relevant?* New York: Kluwer Academic/Plenum.

Yates, J. F., Lee, J., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes, 65,* 138–147.

Zechmeister, J. S., & Zechmeister, E. B. (2000). Introductory textbooks and psychology's core concepts. *Teaching of Psychology, 27,* 6–11.

Zill, N., & Winglee, M. (1990). *Who reads literature?* Cabin John, MD: Seven Locks Press.

Zimbardo, P. G. (2004). Does psychology make a significant difference in our lives? *American Psychologist, 59,* 339–351.

Zweig, J. (2001, May). Do you sabotage yourself? *Money Magazine,* pp. 75–78.

Zweig, J. (2002, October). Are you wired for wealth? *Money Magazine,* pp. 75–83.

心理学教材中译本系列

心理学与生活（第 16 版），菲利普·津巴多、理查德·格里格 著，王垒 等译
教育心理学（第 7 版），罗伯特·斯莱文 著，姚梅林 等译
社会心理学（第 8 版），戴维·迈尔斯 著，侯玉波、乐国安、张智勇 等译
组织行为学（第 11 版），弗雷德·鲁森斯 著，王垒 等译
人力资源管理（第 7 版），劳埃德·拜厄斯、莱斯利·鲁 著，李业昆 等译
人力资源管理（第 10 版），韦恩·蒙迪 著，谢晓非 等译
异常与临床心理学，保罗·贝内特 著，陈传锋、严建雯、金一波 等译
理解孩子的成长（第 4 版），彼得·史密斯 等著，寇彧 等译
心理学（第 7 版），戴维·迈尔斯 著，黄希庭 等译
健康心理学（第 3 版），简·奥格登 著，严建雯、陈传锋、金一波 等译
自我，乔纳森·布朗 著，陈浩莺、薛贵、曾盼盼 译
决策与判断，斯科特·普劳斯 著，施俊琦、王星 译
亲密关系（第 3 版），莎伦·布雷姆 等著，郭辉、肖斌、刘煜 译
态度改变与社会影响，菲利普·津巴多、迈克尔·利佩 著，邓羽、肖莉、唐小艳 译
影响力心理学，菲利普·津巴多、迈克尔·利佩 著，邓羽、肖莉、唐小艳 译
管理决策中的判断（第 6 版），马克斯·巴泽曼 著，杜伟宇、李同吉 译
阅读障碍与阅读困难——给教师的解释，达斯 著，张厚粲、徐建平、孟祥芝 译
APA 出版手册（简明版），美国心理学会 编著，周晓林、叶铮、张旋、曹琳 译
心理学与我们（第 7 版），罗伯特·费尔德曼、黄希庭 著，黄希庭 等译
心理学实验的设计与报告（第 2 版），彼得·哈里斯 著，吴艳红 译
心理学研究方法（第 7 版），约翰·肖内西 著，张明、吴艳红、郭秀艳 等译
危机中的青少年（第 3 版），麦克沃特 等著，寇彧 等译
心理学精要（第 5 版），戴维·迈尔斯 著，黄希庭 等译
跨文化社会心理学，史密斯 等著，严文华 等译
心理统计导论（第 9 版），理查德·鲁尼恩 等著，林丰勋 译
改变心理学的 40 项研究（第 5 版），罗杰·霍克 著，白学军 译
像心理学家一样思考（第 2 版），唐纳德·麦克伯尼 著，王伟平 译
亲密关系（第 5 版），罗兰·米勒 等著，王伟平 译

心理学史（第 4 版），戴维·霍瑟萨尔 著，郭本禹、魏宏波、朱兴国、王申连 等译

人格心理学（第 2 版），兰迪·拉森、戴维·巴斯 著，郭永玉 等译

生物心理学（第 10 版），詹姆斯·卡拉特 著，苏彦捷 等译，彩印精装

孩子的世界：0~3 岁（第 11 版），黛安娜·帕帕拉 等著，陈福美 等译，彩印精装

心理科学之门，阿曼达·阿尔本 著，徐展 译

对“伪心理学”说不（第 8 版），基思·斯坦诺维奇 著，窦东徽、刘肖岑 译

50 位最伟大的心理学思想家，诺埃尔·希伊 著，郭本禹、方红 译

认知心理学及其启示（第 7 版），约翰·安德森 著，秦裕林、程瑶、周海燕、徐玥 译

心理学英文影印版系列

（教育部高等学校心理学教学指导委员会推荐用书）

心理学与生活（第18版），理查德·格里格、菲利普·津巴多 著

普通心理学（第6版），罗伯特·费尔德曼 著，黄希庭教授推荐

心理学实验的设计与报告（第2版），彼得·哈里斯 著，沈模卫教授推荐

心理统计（第9版），理查德·鲁尼恩 等著，张厚粲教授推荐

心理学研究方法（第6版），约翰·肖内西 等著，周晓林教授推荐

社会心理学（第8版），戴维·迈尔斯 著，彭凯平教授推荐

发展心理学（第9版），黛安娜·帕帕拉 等著，林崇德教授推荐

变态心理学（第9版），劳伦·阿洛伊 等著，王登峰教授推荐

心理测验与评估（第6版），罗纳德·科恩 等著，彭凯平教授推荐

认知心理学基础（第7版，双语版），里德·亨特 等著，傅小兰教授推荐

心理测验与评估学习指南，罗纳德·科恩 著

心理统计学习指南（双语版），戴维·皮滕杰 著，林丰勋教授译注

异常与临床心理学，保罗·贝内特 著，陈传锋教授推荐

人格心理学（第2版，双语版），兰迪·拉森 等著，郭永玉教授推荐

生物心理学（第9版），詹姆斯·卡拉特 著，苏彦捷教授推荐

心理测验：历史、原理及应用（第5版），罗伯特·格雷戈里 著，闫巩固教授推荐

英汉对照心理学大词典（第3版），阿瑟·雷伯 等著，王垒 等译注

变态心理学（第13版），詹姆斯·布彻 等著，贾晓明教授推荐

社会研究方法（第6版），威廉·纽曼 著，辛涛教授推荐

如何成为质性研究专家（第3版），科琳·格莱斯 著，刘力教授推荐

改变心理学的40项研究（第5版），罗杰·霍克 著，白学军教授推荐

这才是心理学（第8版），基思·斯坦诺维奇 著，杨中芳教授推荐

当代组织行为学，莉·汤普森 著，谢晓非教授推荐

教育心理学（第7版，双语版），罗伯特·斯莱文 著，姚梅林 等译注